A COMPARISON BETWEEN PERT DISTRIBUTION AND SEASONAL ARIMA MODEL TO FORECAST RAINFALL PATTERN

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yan Li

May 2018

ABSTRACT

Weather is becoming more and more unpredictable for farmers, and the frequency of extreme weather events is also increasing. Small households in Kenya are vulnerable to these extreme weather shocks, and failures in effective hedging will make sustainable production extremely difficult for them. The goal of this thesis is to use historical rainfall record in Kenya to forecast rainfall and take quantile of the rainfall distribution to get a trigger for a put-option embedded innovative financial instrument. There are two methods to develop this lower 20% band trigger, which are pert distribution and time series method. Finally, I get two sets of results from two methods. With the help of the simulation results, insurance companies will be able to design a weather-index insurance for small households in Kenya. For farmers, they will use this flexible insurance as an effective substitute for traditional deposit, which requires productive assets as collaterals.

BIOGRAPHICAL SKETCH

Yan Li is currently a Master of Science student at Cornell University, Dyson School, with a major in applied economics. Her concentration is on agricultural finance.

To my family, friends, all kind minds who used to help, darkness, sorrow, courage and Hope.

ACKNOWLEDGMENTS

I would like to express my deep gratitude to Professor Turvey for the patient guidance, support and suggestions. He encourages me to try harder, which stretches my limit a lot. I would also like to thank Professor Woodard and Professor Shee for their advice on the methodology and structure of the thesis. In addition, I would like to thank all professors whom I had classes of at Cornell University, and without their guidance, I can't finish this thesis, neither.

I would like to thank all my friends, and I know without your help and support, I can't finish this, neither.

Finally, I want to thank my mon and dad. Dad, thanks so much for never giving up on me. You're the kindest person that I've ever known, and you taught me on how to be a good person even under a bad situation. Mom, hope you can always be in a good mood and be healthy. Thanks so much for everything. I want to share this happiness with you.

LIST OF TABLES

Table 3.1: LR Plot 208, Central Machakos: Simulation--50

Table 4.1: SARIMA Model Parameters and Coefficient--73

Table 4.2: Cumulative Rainfall Trigger table (mm) --78

LIST OF FIGURES

Figure 2.1: Risk Contingent Credit Illustration--17

Figure 2.2: Map of Machakos--23

Figure 2.3: Food Expense Percentage Among Total Budget---26

Figure 3.1: Monthly Average Rainfall Results Across 35 years---------------------------------------43

Figure 3.2: Time Series Plot of Monthly Rainfall for Example Area Across 1981 to 2015--------44

Figure 3.3: Time Series Plot of Monthly Rainfall for Example Area Across 1981 to 2015 with Trend line--45

Figure 3.4: Long Rain Correlation Matrix---47

Figure 3.5: Short Rain Correlation Matrix---47

Figure 3.6: Long Rain and Short Rain Combination Correlation Matrix----------------------------48

Figure 3.7: Long Rain & Short Rain Overlapping, Central Machakos-------------------------------49

Figure 3.8: Pert Distribution Simulation---51

Figure 3.9: Dickey-Fuller Distribution--56

Figure 3.10: Seasonal Differencing Distribution, Central Machakos--------------------------------58

Figure 3.11: Cumulative Distribution of Seasonal Differencing, Central Machakos---------------59

Figure 3.12: Seasonal Differencing Distribution, Central Yathui-------------------------------------60

Figure 3.13: Cumulative Distribution of Seasonal Differencing, Yathui------------------------------60

Figure 3.14: Seasonal Differencing Distribution, Yatta---61

Figure 3.15: Cumulative Distribution of Seasonal Differencing, Yatta-------------------------------61

Figure 3.16: Seasonal Differencing Distribution, Masinga--62

Figure 3.17: Cumulative Distribution of Seasonal Differencing, Masinga---------------------------62

Figure 3.18: Seasonal Differencing Distribution, Matungulu---63

Figure 3.19: Cumulative Distribution of Seasonal Differencing, Matungulu------------------------63

Figure 3.20: Seasonal Differencing Distribution, Kalama--64

Figure 3.21: Cumulative Distribution of Seasonal Differencing, Kalama---------------------------64

Figure 3.22: Seasonal Differencing Distribution, Kathiani---65

Figure 3.23: Cumulative Distribution of Seasonal Differencing, Kathiani--------------------------65

Figure 3.24: Seasonal Differencing Distribution, Mwala---66

Figure 3.25: Cumulative Distribution of Seasonal Differencing, Mwala----------------------------66

Figure 3.26: Seasonal Differencing Distribution, Kungundo--67

Figure 3.27: Cumulative Distribution of Seasonal Differencing, Kungundo------------------------67

Figure 3.28: Seasonal Differencing Distribution, Ndithini--68

Figure 3.29: Cumulative Distribution of Seasonal Differencing, Ndithini---------------------------68

Figure 3.30: Seasonal Differencing Distribution, Mavoko--69

Figure 3.31: Cumulative Distribution of Seasonal Differencing, Mavoko---------------------------69

Figure 3.32: ACF & PACF Plot of Original Data---70

Figure 4.1: Plot of Residuals--74

Figure 4.2: ACF & PACF of Residuals---75

Figure 4.3: Distribution of Simulation and 20% lower quantile, Central Machakos---------------77

Figure 4.4: All Triggers Comparison Between Two Methods---79

Figure 4.5: Trigger Comparison Between Two Methods, Central Machakos-----------------------80

Figure 4.6: Trigger Comparison Between Two Methods, Kalama-----------------------------------80

Figure 4.7: Trigger Comparison Between Two Methods, Kangundo---------------------------------81

Figure 4.8: Trigger Comparison Between Two Methods, Kathiani-----------------------------------81

Figure 4.9: Trigger Comparison Between Two Methods, Masinga-----------------------------------82

Figure 4.10: Trigger Comparison Between Two Methods, Matungulu--------------------------------82

Figure 4.11: Trigger Comparison Between Two Methods, Mavoko-----------------------------------83

Figure 4.12: Trigger Comparison Between Two Methods, Mwala-------------------------------------83

Figure 4.13: Trigger Comparison Between Two Methods, Ndithini-----------------------------------84

Figure 4.14: Trigger Comparison Between Two Methods, Yathui-------------------------------------84

Figure 4.15: Trigger Comparison Between Two Methods, Yatta---------------------------------------85

TABLE OF CONTENTS

Chapter 1: Introduction--1

1.1: Introduction--1

1.2: Economic Problem--3

1.2.1: Background--4

1.2.2: Risk Contingent Credit--5

1.2.3: Credit-rationing--7

1.2.4: Poverty trap---8

1.2.5: Basis Risk---9

1.2.6: Covariate risk--10

1.3. Research Problem---11

1.4: Overall objective and secondary objective---12

1.5: Roadmap of thesis--12

1.6: Summary of thesis--13

Chapter 2: Risk Contingent Credit Review and the Kenya Pilot Project----------------------------14

2.1: Introduction--14

2.1.1: Risk Transfer Method--15

2.2: Background of RCC---16

2.2.1: Interest Premium Formula--17

2.3: Overview of Machakos County Pilot Program--22

2.3.1: Some Observations from Pilot Program---24

2.4: Other Consideration in Design RCC--27

2.4.1: Internal Risk Management Method---27

2.4.2: External Risk Management Method--27

2.4.3: Basis Risk--29

2.4.3.1: Local basis risk---29

2.4.3.2: Geographic basis risk---30

2.4.3.3: Product basis risk--30

2.5: Related Program---31

2.5.1: Direct Cash Payment Method--31

2.5.2: IBLI--33

2.6: Lessons---37

Chapter 3: Data Source, Characteristics and Time Series Method ------------------------------42

3.1: Introduction--42

3.2: Model in pilot: Pert Distribution---45

3.3: Correlation and Covariate Risk--46

3.4: Data Transformation--51

3.4.1: Stationarity--53

3.4.2: Procedure---54

3.4.3: Overfitting--55

3.4.4: Unit Root- ---55

3.4.5: Rainfall Differencing Operation---57

Chapter 4: Results--71

4.1: Seasonal ARIMA model and Parameters selection------------------------------71

4.1.1: Model Selection Criteria--71

4.1.2: Application---73

4.1.3: Residual Check--74

4.1.4: Forecast and lower 20% quantile of simulation-----------------------------------76

4.2: Results Comparison---78

Chapter 5: Conclusion---86

Reference---91

Appendix---95

Chapter 1: Introduction

1.1 Introduction

The most significant risk to farmers in Sub-Saharan Africa is the failure of periodic rains. Periodic rains are rainfalls happening between October and January, known as "long rains", and between March and May, known as "short rains". When either season of periodic rain fails, farmers face substantial hardship from the loss of their crops. These crop failures lead to loss of livelihoods, reduced expenditures on food and education, and the sale of productive assets, which leads farmers into a poverty trap that is difficult to escape without external help (Makaudze, E. 2012; Kumar, Turvey, and Kropp, 2013). This thesis explores insurance solutions to compensate farmers for losses during adverse weather events, and a credit pricing issue that is related to a pilot program for risk contingent credit in Kenya.

A study conducted by Green Revolution in Africa (AGRA) reports that farmers in Sub-Saharan Africa have recently experienced more unpredictable weather patterns than they have in the past (Makaudze, E. 2012). The frequency of extreme weather events, especially droughts in Africa, is increasing significantly, and farmers, especially small households, are becoming more vulnerable to unpredictable weather patterns. Given such uncertainty about weather conditions, farmers would benefit greatly from an innovative financial instrument to help them hedge against losses from adverse weather events. However, farmers face a market failure in this regard; they lack any sort of insurance that would protect them against losses from losing their crops. Further, financial institutions are unwilling to provide credit to small household farmers, due to their high default rates. In the absence of affordable credit, farmers don't have enough capital to purchase inputs, expand capacity, or to implement new technology.

Smallholder farmers need flexible and accessible credit solutions, which can help them hedge volatility in production caused by unpredictable weather shocks. Risk contingent-credit (RCC) is one of the instruments designed to properly guide agricultural production (Shee, Turvey, Woodard, 2015; Shee, Turvey 2012). A pilot program based on RCC is currently underway in Kenya. The pilot program of RCC is in Machakos County, which is a maize growing area, with some intercropping with perennial fruits or other cash crops (Shee, Turvey, You, 2015). Most farmers in this area are small households who are vulnerable to extreme weather event. There are two major rainfall seasons, which are long rain season and short rain season. Those two rainfall seasons contribute to the major rainfall of the area. This RCC pilot program want to set a dynamic mechanism in credit, which connects loan repayment to the performance of underlying asset, which is rainfall here to prevent farmers from droughts. Finally, farmers can maintain a stable cash flow either during the good rainfall season and drought.

In some other countries, such as Mexico, weather index insurance was approved by the government and emerged out of a similar pilot program (Yucemen, 2005) applied the mechanism as RCC pilot program in Kenya. However, in Africa, the application of this insurance is still at its initial stage. Some other RCC-related practices pay attention to spatial characteristics, such as temperature and precipitation, to construct models for RCC. For example, a cooling degree day (CDD) index is used as a benchmark to quantify spatial risks brought by weather (Norton, Turvey, Osgood, 2012). In the case of Kenya, rainfall will be the weather event of most importance. Nowadays, institutions ranging from local governments to international organizations are paying attention to the initiation of this innovative financial instrument. Of course, there is an enormous potential demand among farmers, especially those poor small household farmers trapped in poverty.

As local insurers lack experience with weather-index insurance policies, the pricing process needs to be better understood before the market for such weather-index insurance ultimately attains the desired breadth and depth. In addition, local governments don’t have relevant experience, and they will also need more information to be able to customize the policies and regulations related to weather-index insurance.

In the Kenya pilot project, rainfall triggers were based on Monte Carlo Simulation of cumulative rainfall, which follows an empirical PERT distribution. To advance principal of efficiency and equity in RCC design, it is important to consider alternative risk metrics. This thesis tries to use an alternative time series methods to model rainfall patterns in Kenya.

1.2 Economic Problem

The economic problem investigated throughout this thesis is market failure in the insurance and credit markets in Kenya. In the absence of a marketable collateral, insurance can provide an effective substitute for collateral (Shee, Turvey, 2012). With the existence of insurance, small famers could theoretically access credit, while retaining valuable productive collateral. The preservation of collateral could solve credit-rationing problems, and farmers would be more willing to use credit, while lowering risk in production and improving their quality of life. In contrast to the current situation in Kenya where small farmers must seek external support to hedge themselves from weather shocks, RCC can provide farmers with a comparatively inexpensive way to protect their productive assets, either partially or fully.

If rainfall patterns can't be captured accurately and covariate risks can't solved smoothly, it will be difficult to lend money to local farmers from a risk management perspective. Incorporating insurance into a credit product will make lending more accessible by lowering covariate risks. Finally, lending to small households in Kenya will be better positioned, and credit supply in agriculture will be increased accordingly.

1.2.1 Background

The problem of credit in developing countries impedes agricultural productivity and therefore causes poverty trap potentially. Uncertainties in weather and poor local infrastructure construction affect credit use, and marketing opportunities in agriculture. Those adverse situations force farmers to use low-risk and low-return activities during the production, and finally make poverty a wide-spread and persistent problem in developing countries (Shee, Turvey, Woodard, 2015).

Credit constraint is a significant issue, which prevents farmer from escaping poverty. Credit-constrained households refer to conditions: 1. Farmers won't enter into credit market, because they may assume that they will be denied credit; 2. After applying the loan, farmers are refused. 3. Farmers receive less amount of loan than their previously requested amount. Credit-constraints are significant problems in developing countries, affecting the efficiency of farmers' production in agriculture. Especially, farmers with lower-asset bases are more significantly affected by credit constraints than those in higher-asset classes (Kumar, Turvey, Kropp, 2013).

The target group of this RCC is small households in developing countries. This group of farmers will be significantly influenced by extreme weather events, and meanwhile, their situations can be changed positively by applying credit-enhancing instruments, such as RCC during the production.

1.2.2 Risk Contingent Credit

Weather insurance can help small farmers access the capital market by using RCC-embedded financial instrument provided by the counterparties. For those counterparties, RCC-embedded financial instrument provides those players an opportunity to initiate a new business and generate profit. With access to RCC, farmers can maintain a higher production level, and ultimately use new technology in agricultural production to increase production efficiency and hedge against negative weather shocks. Given the objectivity of credit-linked characteristics, RCC-embedded financial products can align interests for both farmers and insurers, which solves the agency problem. Compared with traditional claim-based insurance, new index-based insurance can shape the farmers' behaviors to avoid fraud, adverse selection and moral hazard, since index insurance payouts depends on measures of an external benchmark, and not on case-by-case assessments of individual's loss. This objective characteristic cannot be manipulated and therefore, it guarantees the accuracy of the insurance. Offering agricultural insurance and agricultural credit has potential to raise an agency problem, which indicates conflict of interest between small households and insurance issuers under current situation in agricultural insurance.

Linked-credit products or risk contingent credit is a good solution to this problem. In addition, the fact that most small farmers live in remote areas in Kenya increases the cost of traditional claim-based insurance. This new insurance can decrease the transaction cost, because agents will have different means of information verification, and payout determination, which is much cheaper than traditional claim-based insurance. For example, poor infrastructure construction, such as road transportation, will make it expensive for appraisers to go to the field to examine the real loss, while the new insurance can simply use rainfall measurements collected by weather stations, recorded remotely.

The effectiveness of credit-linked insurance would depend heavily on how well basis risks and covariate risks are accounted for. Any model with poor quality results won't solve misbehavior problems significantly, and farmers won't benefit from the insurance. Also, corresponding laws, regulations, and infrastructure are also considerations in designing credit-linked insurance. For example, when pricing the insurance, the research team will refer to historical data and apply innovative technologies, such as remote sensors, to create an accurate benchmark. A complete infrastructure for measuring rainfall and insuring against rain failures would ensure that farmers secure a stable income, which comes from the sales of their agricultural product, insurance repayment, or both.

This thesis seeks to find an accurate way to price weather-index insurance, so that basis risk can be reduced, and actual rainfall pattern can be implemented into insurance pricing mechanism. Rainfall is the only factor that affects interest premium. More parameters, such as temperature, and soil moisture, will be discussed later, as well as the correlation among these factors.

1.2.3 Credit-Rationing

Default risk and the size of potential losses are two main factors used to assess credit risk. High default probabilities of small farmers in Kenya makes traditional agricultural insurance unprofitable. In addition, fear of losing their collateral, which is a typical example of risk-rationing (Boucher, S.R., Carter, M.R., Guirkinger, C. 2008; Verteramo-Chiu, Khantachavana, Turvey, 2014), also drives farmers away from traditional insurance products (Stiglitz, Weiss, 1981).

The definition of risk rationing is officially put forward by Boucher (2008):" Risk rationing occurs when insurance markets are absent, and lenders, constrained by asymmetric information, shift so much contractual risk to the borrower that the borrower voluntarily withdraws from the credit market even when he had the collateral wealth needed to qualify for a loan contract."

Farmers' risk rationing behaviors in other developing countries, such as China and Mexico were intensively investigated previously (Verteramo-Chiu, Khantachavana, Turvey, 2014). Results turn out to be that risk rationing does exist among farmers in China and Mexico. Therefore, risk enhancing activities supported by RCC will significantly change farmers behaviors, if RCC product can make lending more accessible and inexpensive to farmers (Verteramo-Chiu, Khantachavana, Turvey, 2014).

In a traditional insurance market, farmers are required to deposit productive assets as collateral in exchange for credit. However, in the case of weather-index insurance, farmers only pay a premium which represents an mean of the expected loss, and hence farmers don't have to be afraid of losing productive collateral due to extreme weather events. However, during the repayment of the insurance, there are type I error and type II error about basis risk. For example, type II indicates that there is drought and farmers' production is negatively affected, however, the insurance is not triggered. Similar situation may happen on type I error. It indicates that there is no drought, however, the RCC product is exercised and farmers get paid, even though they don't suffer from the drought.

An efficient weather index-insurance market would make credit markets more accessible to small farmers and therefore contribute to increase in productivity, as farmers would be able to free more resources to buy machinery, higher quality seed and fertilizer, and more technologically advanced equipment that would aid in protecting their crops from increasingly volatile weather changes. Of course, risks in different areas of Kenya are not the same, and so the actual premium must be modified accordingly to reflect different risks associated with different regions.

1.2.4 Poverty Trap

Adverse weather shocks put farmers in Kenya into a poverty trap. Small households can be easily put into poverty trap due to vulnerability to weather shocks, if they can't have a stable cash flow. However, with the help of RCC, farmers can maintain a stable cash flow in either good or bad weather conditions. For example, during the harvest time, farmers can generate cash flow by selling crop; however, during the drought season, farmers can get compensated by RCC product accordingly.

Poverty indicates that farmers can't generate income which is sufficient to cover their basic consumption needs. There are many reasons that cause poverty. First, farmers own few assets, or the assets are of a low quality. Second, households don't have sufficient access to technology, capital or credit due to poor local infrastructure or lack of required knowledge. Third, poor enforcement of laws and regulations will also contribute to poverty. A poverty trap is a dynamic equilibrium, tied to the agricultural production levels for households (Barnett, Skees, 2008). If the level of production is above a certain threshold, production will increase to a high-output equilibrium. If the level of production is below that threshold, production will diminish to a low-output equilibrium below the poverty line, which means that farmers could fall into a poverty trap.

Farmers with different levels of wealth will have different losses due to the same weather shocks. For example, wealthy farmers are likely to have more advanced technologies that reduce the losses caused by weather. However, small households, recovering from weather shocks, will face greater difficulties than wealthy farmers due to their lack of resources. RCC-embedded instruments, which facilitate the connections between farmers and public financial markets, provides a feasible solution for this complicated problem.

1.2.5 Basis Risk

One major risk of a weather-based index is basis risk, which indicates that the risk that payoffs of a hedging instrument don't correspond to the underlying exposure (Norton, Turvey, Osgood, 2012). Basis risk creates two issues for agricultural insurance: on the one hand, even if farmers suffer from extreme weather events, they don't get paid because the trigger of the insurance wasn't exercised. On the other hand, farmers will receive indemnity due to the mismatch between the trigger and actual rainfall, and the amount of the indemnity is higher than the actual loss caused by the adverse events. High basis risk undermines the willingness of potential customers to purchase weather insurance.

Due to poor information infrastructure in Africa, accurate data is scarce, and limited data availability affects the accuracy of the model. The sample must be large enough to draw reliable results, but many data are not available, which poses a challenging problem that needs to be overcome in order for accurate pricing models to be determined. Without enough data to estimate the model, the efficiency of the insurance is in jeopardy. For example, rainfall record is a significant factor in RCC product pricing procedure, and data set we used here is provided by CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data). The interval between two observations is 10 days, and it ranges from 1980 to 2016.

1.2.6 Covariate Risk

In designing the insurance, covariate risk, also known as systematic risk, will significantly influence the performance of the insurance. For weather-index insurance, covariate risk refers to the tendency of extreme weather events to influence an area and its neighboring areas simultaneously. Systematic risk will potentially cause insurance companies to suffer from huge losses, thus making such weather-index insurance less attractive for them to design. Insurance companies can’t avoid systematic risk within their portfolios, because almost all farmers in a given area will encounter similar weather shocks, meaning they cannot effectively diversify against this risk at a local level. In the case of extreme events, such as droughts, spatial correlations will be more obvious, as a large area will be effected contemporaneously. Nonetheless, insurance companies can diversify systematic risks within their international portfolio, by insuring across many regions with different weather patterns.

1.3 Research Problem

The research problem lies in designing a fairly-priced financial instrument. There are two key risks related to designing the model for insurance, namely the covariate risks and basis risks. For RCC to be effective, it must be determined how to price the insurance, and this needs to be established on an actuarial basis.

The current pilot program, implemented in September 2017, bases insurance calculations on the Pert Distribution, which is derived from the historical rainfall data measured by remote sensors. An alternative approach is to use ARIMA processes to measure the forward risks and price the insurance accordingly. Output generated from ARIMA model will be compared with the output from Pert Distribution, which is currently being employed. SARIMA model is used here as an alternative method because we want build a model for equally-spaced observations with periodic pattern. With the help SARIMA model, we can either capture the connection from both the previous month and also the months one year ago.

1.4 Overall objective and secondary objective

The aim of this thesis is to create a model to simulate rainfall and generate input for more accurate insurance pricing. In this regard, there are two methods, namely the pert distribution model and time series ARIMA model. A trigger, which is a regression result generated by historical rainfall data, will be selected according to the simulation results.

Currently, the pert distribution method is being applied in pilot program in Kenya. An ARIMA model will be introduced here to generate another possible input for the pricing of the insurance. Results from two different methods will be compared, and the better one will be selected as the input for the insurance pricing model.

1.5 Roadmap of Thesis

In this paper, pert distribution and time series model will be applied to simulate the pattern of rainfall. The rainfall data set, provided by the Kenyan government, contains bi-weekly rainfall records from ten locations in Kenya, over a span of 35 years. One location has been chosen as an example to apply rainfall simulation.

For the pert distribution, risk will determine the parameters of model. For the time series analysis, seasonal ARIMA (Auto-Regressive Integrated Moving Average) will be applied to capture rainfall characteristics. With the results of these two models, a benchmark for the trigger of this index insurance can be created.

With this weather-index credit solution, smallholder farmers can pay an affordable interest premium that is connected to the underlying drought risk, and deal with the rainfall uncertainty more easily in their production decisions. There is a tradeoff between the interest rate charged on the loan, which has an embedded option, and the collateral requirement for the loan. For farmers with limited collaterals, or who fear losing their collaterals, the insurance is an effective way to hedge risks, and it is an affordable substitute for providing collateral (Carter, 2007).

Most importantly, risk-contingent credit is a financial product that deals with the downside risk, especially drought, and this index insurance can also stimulate both the supply and the demand of the insurance product.

1.6 Summary of Thesis

The second chapter is a literature review that provides more details about the background and the economic theories of the paper. The third chapter introduces the pert distribution method and the time series method. The fourth chapter examines results generated by these two models, and the fifth has the conclusion, along with directions for future research.

Chapter 2: Risk Contingent Credit Review and the Kenya Pilot Project

2.1 Introduction

Kenya lies in the east of Africa, and has high temperatures and low rainfall. Small household farmers in Kenya are vulnerable to weather shocks, and there is a need for a new financial instrument to hedge against weather risks. Traditional insurance can widen farmers' access to credit markets, but it has its own shortcomings. (1). Moral hazard and asymmetric information are inevitable in traditional insurance environments. (2). Loss verification based on individual claims can be costly, especially for remote areas. (3). Traditional insurance companies will be reluctant to lend money to those risky small farmers, due to high default probability, which is a risk rationing issue and it will be discussed in detail later; however, there are needs existing among farmers. Even though small households are willing to pay a higher premium to purchase the insurance, companies still have their concerns about the cost of those transactions, due to farmers' vulnerability to extreme weather events. (4). Poor transportation networks impede small farmers from using capital markets physically. (5). Subsidies are controversial, and can be unaffordable for local governments. On the one hand, subsidy from local government can reduce the cost of issuance of the insurance immediately and therefore attract more insurance companies to participate in the provision of the insurance; on the other hand, direct subsidy may distort the market structure. It distorts the relationship between supply and demand. If external subsidy is removed from the market, the demand for the insurance may collapse significantly (Makaudze, E. 2012). These are the real problems farmers in Kenya currently face. This chapter will provide more detailed background information about pilot program in Kenya and weather-index insurance.

2.1.1 Risk Transfer Methods

There are many ways to transfer risk. For example, "Semi-formal microfinance and socially-constructed reciprocity obligations within village, family, religious community are informal ways to do risk management" (Coate & Ravallion, 1993; Fafchamps & Lund, 2003; Grimard, 1997; Rosenzweig, 1988; Townsend, 1994, 1995).

Informal microfinance mainly focuses on small households directly. Semi-formal microfinance brings local cooperative organizations, government, and international non-profit organizations into the effort to help small household farms.

Besides the traditional methods mentioned above, RCC is an innovative financial product, which provides a new and flexible way to make repayment. The repayment mechanism makes it more practical and appropriate for risk management. For small farmers, insurance is an efficient way to transfer risks. For the insurance companies, the risk can then be diversified away across their portfolio of different regions.

In many extremely poor areas, even informal risk transfer methods are not widely available. If farmers are well-funded like many wealthy peers, they can use their own income to purchase additional assets to hedge and to improve the efficiency of production. However, in Africa, capital flow is impeded by high transaction cost due to the lack of risk transfer mechanism. It will be expensive for farmers to detect potentially possible ways to borrow money, due to the lack of effective information platform. Meanwhile, lack of effective information platform is also a problem for insurance company to find potential clients who have demands in the external help to hedge themselves against drought.

A risk transfer mechanism can be influenced by many external factors, such as risk aversion, financial liquidity, understanding of the product, trust in the provider and access to the market (Jensen, Mude, Barrett, 2014). The characteristics of target clients and the local infrastructure will influence the effectiveness of risk management, which requires that researchers not focus solely on the insurance and consider other external factors such as the size of the loan, timing of the insurance, and repayment function of the insurance, when they try to apply this risk transfer product locally.

2.2 Background of RCC

Risk contingent credit is a risk management method, which connects the payoffs of the loan and the performance of underlying asset, which is rainfall here. RCC uses underlying indicators, such as rainfall here to quantify the level of drought. The RCC can attract risk-rationed farmers to use credit, because RCC removes the possibility of losing all collaterals caused by the failure of paying back the loan on time. Finally, RCC can transfer risk from farmers, who are borrowers of the credit to the lender by getting access to credit market and using agricultural insurance. Compared to traditional credit tools which required collateral, RCC is more flexible for farmers, because the cost of using RCC is the premium charged on the loan, which is more affordable than the potential loss the whole productive asset.

Figure 2.1: Risk Contingent Credit Illustration

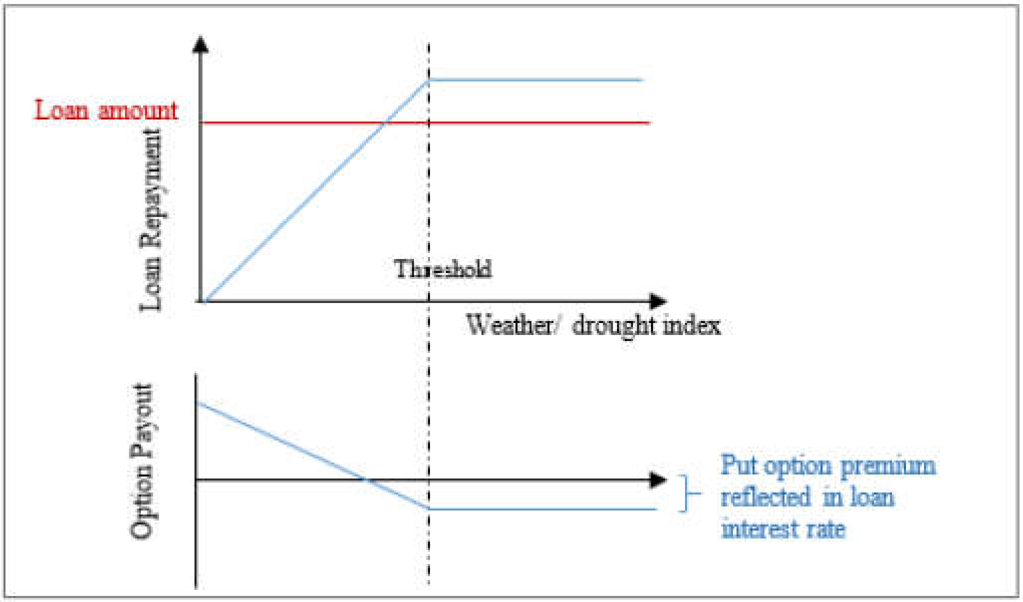


Figure 2.1 explains the mechanism of RCC. RCC requires farmers to pay risk premium, which is incorporated with local risk factor, and it insures against drought. The upper graph shows that the amount of loan repayment is related to drought. Before the threshold, which is the numeric benchmark defined drought, loan repayment is positively correlated to rainfall. The lower graph indicates the payout function of the loan, which is negatively correlated with rainfall. It shows that if rainfall goes above the trigger, there is no drought in the area, and technically speaking, farmers can have a sustainable product which can generate enough money to support the production in the following season.

2.2.1 Interest Premium Formula

The input trigger, which is a quantile threshold, is generated through simulation, and will be put into premium calculation equation. If actual rainfall falls below the trigger, then the insurance will be executed to repay the loan. In the following function, rainfall simulation result will be converted into an interest rate result, which includes the insurance premium.

$$i^* = \frac{\ln[\frac{\psi E\left[MAX\left(0, Z - P(T)\right)\right]}{f} + e^{i**T}]}{T}$$

(Shee, 2015 and Turvey, 2012)

The variables in the above equation are defined as follows.

T: Time length. Here T=1, which is 1 year.

ψ: Hedge ratio. A leverage ratio, which indicates farmers don't take any financial leverage to multiply their losses or profits when it is equal to 1.

f=Z, which is the trigger of this put option-embedded instrument, given that the hedge ratio equals 1. Because the insurance will compensate for the drought, which means that actual rainfall is smaller than estimated , we can consider this insurance as a put option, because put option will be exercised only when the underlying asset price (actual rainfall) is lower than the strike price (input trigger).

$E\left[MAX\left(0, Z - P(T)\right)\right]$: Average return of this option, and the unit is in mm (rainfall). Z is the trigger level for the local rainfall, and P(T) is the actual local rainfall.

The expected rainfall payoffs are different for different plots. Even for same plot, the trigger will be different between long rain season and short rain season. Under the current situation, the product is designed to be a put-option embedded instrument, and it can also be designed as a call option, which is designed to compensate for a flood.

i**: base interest rate. It will generally be adjusted according to the local interest rate, and here the default value is 12%.

After plugging in all these parameters, new interest rate will be calculated. Furthermore, we assume that there is a 25% profit margin for the insurance company. The final premium should be the original premium plus a 25% profit margin.

Finally, insurance provides a flexible option for farmers, and it is a substitute for a traditional collateral deposit. This method reduces farmers' debt obligations (Shee, Turvey, 2012), because it removes the pre-requirement of collateral, and it provides a dynamic way to connect environmental factors and actual underlying production.

The equation above is a general formula used to calculate the premium. Next, a more accurate version will be provided, which considers the rainfall effects of both long rain and short rain. Ultimately, the formula will provide a concrete interest premium to charge.

The loss of rainfall is calculated as $Max(0, Z_i - R_i(t_i, T_i))$. Z_i is the trigger, which this thesis is going to focus on. $Z_i - R_i(t_i, T_i))$ Indicates the difference between actual rainfall and benchmark. Since the trigger is an objective number based on historical rainfall pattern and can't be manipulated by an individual, the insurance removes the risks of moral hazard. $g_i\left(R_i(t_i, T_i)\right)$ is the probability distribution function of rainfall. Finally, the mean rainfall loss will be calculated as follows:

$$E\left[Max(0, Z_i - R_i(t_i, T_i))\right] = \int_{Min\left(R_i(t_i, T_i)\right)}^{Z_i} \left(Z_i - R_i(t_i, T_i)\right) g_i\left(R_i(t_i, T_i)\right) dR_i \quad (1)$$

According to the function above, the expectation of the rainfall repayment can be written as an integral whose lower bound is the historical lowest rainfall and its upper bound is the insurance trigger. It is the the difference between the trigger and actual rainfall times the probability distribution of the actual rainfall. It compensates farmers only in a drought. If actual rainfall exceeds the trigger value, the payoff of the insurance is 0.

Finally, a tick value is calculated as

$$\psi_i = \frac{f}{Z_i - Min\left(R_i(t_i, T_i)\right)} \quad (2)$$

where f is the loan principal, or the size of the insurance package. $Min\left(R_i(t_i, T_i)\right)$ Indicates the minimum rainfall on record. The tick value is the price of loss in rainfall per mm, which translates rainfall loss into monetary repayment. It is a conversion between lost rainfall (mm) and loan amount (local currency unit).

If actual rainfall falls below the trigger, then farmers will be compensated according to the difference between the benchmark and actual rainfall. Therefore, a severe drought condition will be compensated more than a less severe drought condition, other things constant. If actual rainfall falls below the minimum historical rainfall level, then farmers don't have to repay the loan, and the only cost for them is the insurance premium they paid in advance.

$$B = e^{-iT}\left(fe^{i^*T} - \left(\psi_L\left[Max(0, Z_L - R_L(t_L, T_L))\right] + \psi_S\left[Max(0, Z_S - R_S(t_s, T_s))\right]\right]\right) \quad (3)$$

The equation above indicates the present value of loan repayment after considering rainfall failures in both the short rain and the long rain season. i is the bank's cost of capital. i^{*} Indicates the standard interest rate on operating loans. $fe^{i^{*}T}$ is the future value of principal at a standard interest rate without considering rainfall failures. However, this equation doesn't consider the correlation between the long rain and the short rain. Since there is a positive correlation between them, failure of one rainfall season is a predictor of the failure of another rainfall season. Thus, the equation below tries to capture this kind of correlation.

$$v = \int_{Min(R_L(t_i,T_i))}^{Z_L} \int_{Min(R_S(t_i,T_i))}^{Z_S} \left(\psi_L\left(Z_L - R_L(t_L,T_L)\right) + \psi_S\left(Z_S - R_S(t_s,T_s)\right)\right) g\left(R_L,R_S\right) dR_L dR_S$$

(4)

Moreover, the equation above reflects expected losses of rainfall after considering the rainfall difference between benchmark and actual rainfall for both the long rain season and the short rain season. A joint probability distribution function is applied here to calculate the expected value.

$$E\left[B\right] = e^{-iT}\left(fe^{i^{*}T} - v\right) \quad (5)$$

Therefore, after substituting v into the original equation, a corresponding present value of the operating loan can be concluded from the equation above.

$$B_1 = e^{-iT} fe^{(i^{**})T} \quad (6)$$

This equation calculates the present value of loan without rainfall risk and imbedded insurance. This number is the mean of the index insurance needed to price this product fairly. Especially, for i **, it can be considered as a mean value after considering both the long rain shortage and short rain shortage.

Technically speaking, result from (5) should be equal to result from (6), since equation (5) considers two actual rainfall seasons, and equation (6) is an expected value for the whole season. Therefore, we can equate (5) and (6) and get:

$$e^{-iT}\left(fe^{i^{*}T}-v\right)=e^{-iT}fe^{(i^{**})T} \quad (7)$$

After solving (7), finally we get a modified version of the interest premium:

$$i^{*}=\frac{\ln\left[\frac{v}{f}+e^{(i^{**})T}\right]}{T}.$$

This equation is equivalent to the equation below, after plugging in each individual term. Compared to its original version, this equation distinguishes explicitly between both long rain and short rain.

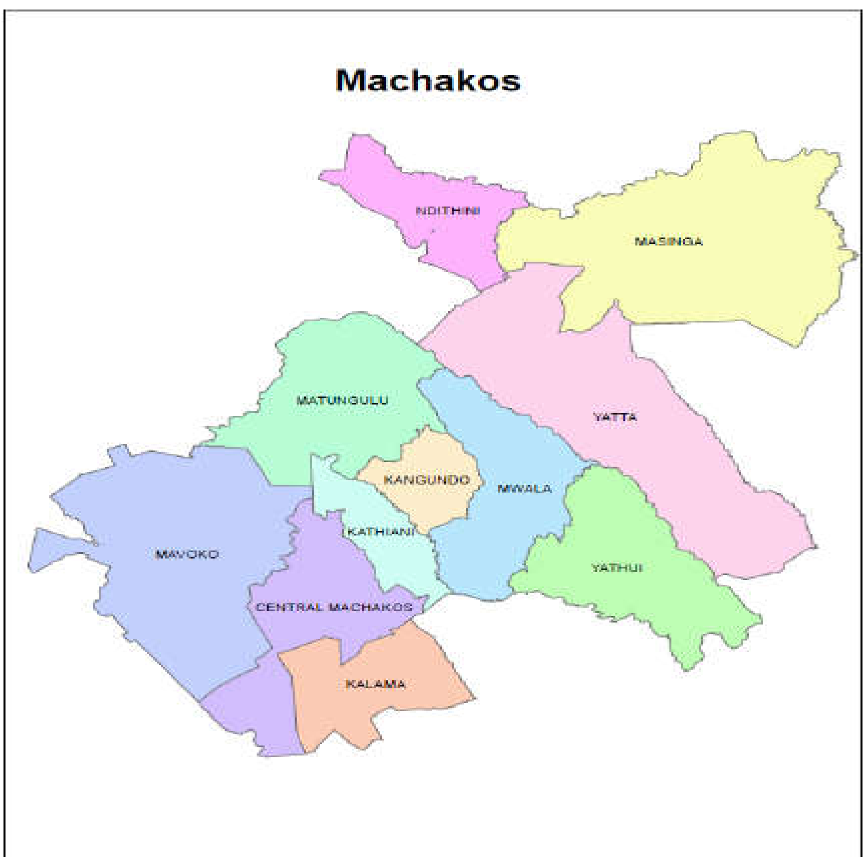
$$i^{*}=\frac{\ln\left[\frac{E\left[\psi_{L}\left(Z_{L}-R_{L}(t_{L},T_{L})\right)+\psi_{S}\left(Z_{S}-R_{S}(t_{s},T_{s})\right)\right]}{f}+e^{(i^{**})T}\right]}{T} \quad (8)$$

2.3 Overview of Machakos County Pilot Project

The current pilot program in Machakos County in Kenya is an application of RCC, aiming to provide alternative methods for small households to do risk transfer. Annual rainfall of this area is low, and farmers are facing production problems brought by drought. Pilot program focuses on rainfall in 11 plots in Kenya. Farmers in Kenya begin the production for the following season by selling crop in the previous season. If the sales of the production don't meet the amount of the previous loan, farmers will fail to pay back the loan directly, due to their low protection level in drought. This cycle is vulnerable to drought, and farmers in Kenya need an efficient alternative method besides traditional loan to finish self-hedging.

Extreme weather events, especially drought, damage agricultural production in Sub-Saharan Africa. Lack of credible credit history makes it risky for an insurance company to issue related financial products to help farmers hedge risks in production. To guarantee the quality and reduce the cost of the insurance, issuers of the insurance tend to impose collateral requirements as a deposit toward the loan in advance, which discourages farmers from applying the insurance. Farmers are credit-rationed on collateral. In a sense, if farmers know there is a possibility of losing their productive asset once they fail to pay back the loan, they may choose to withdraw the loan voluntarily.

Figure 2.2: Map of Machakos



This RCC pilot area covers eleven divisions in the Machakos County including Central Machakos, Yathui, Yatta, Masinga, Matungulu, Kalama, Kathiani, Mwala, Kangundo, Ndithini, and Mavoko, which can be seen from graph above. In those areas, uninsured risks are a major cause of low agricultural productivity, and droughts affect the local agricultural production negatively. With 80% of the population in these areas employed in agriculture and 22% of country's overall GDP derived from agriculture. Therefore, agricultural production in Kenya is a critical issue (Shee, Turvey, You, 2015). Application of RCC product can enhance agricultural production for Kenya, and finally improve the well-being of local population.

Maize is the dominant crop in the area with intercropping with perennial fruits or other cash crops (Shee, Turvey, You, 2015). There are two major rainfall seasons in Kenya. One is the long rain, which begins from October 15th to January 15th, and the other is the short rain, which begins from March 15th to May 15th. There is a positive correlation between two rainfall seasons, and failure in any rainfall season be associated with crop failures in the other season.

2.3.1 Some Observations from Pilot Project

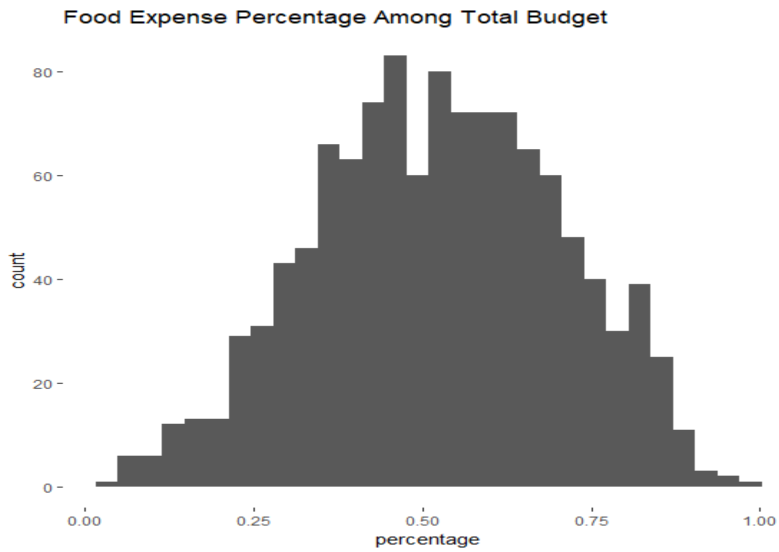
According to survey data collected from pilot program, there are 1166 observations and 33 variables within the dataset. Variables include rationing group, treatment, location, age, education, genders, income, and expenditures. Farmers are assigned into different categories of the insurance, including: RCC, Normal Credit and Control during the pilot. Risk-contingent credit is the innovative financial instrument that this thesis is going to focus on. Normal credit is the traditional loan provided by the local banks or government.

According to the results from the survey, 42% of the households are risk-rationed and voluntarily withdraw themselves from the credit market(Shee, Turvey, You, 2015). Therefore, there is a great demand for credit-insurance product as a substitute for the traditional loan method.

The profiles of farmers are significantly different. For example, the proportion of a farmer's budget spent on food ranges from 4.32% to 99.70%. The mean percentage of food expense among the sum of all expenditures is 27.46%, the median is 27.47% with a variance of 0.0323. The first quantile of the percentage is 13.39%, and third quantile of percentage is 38.09%. More details can be found in graph below.

Wealthy farmers will influence the acceptance rate and the effectiveness of the insurance. For example, for farmers who are at the bottom of the wealth level distribution, weather-index insurance won't be helpful for them, since they cannot afford the premium of the insurance. Meanwhile, for some wealthy farmers, who already have a set of efficient technological methods to hedge against weather risks, they won't find weather index-insurance useful either.

Figure 2.3: Food Expense Percentage Among Total Budget



Source: Kenya Pilot program

42.11% of farmers who were surveyed said that they didn't apply for loan from bank/ cooperative organization, while 28.99% of farmers didn't answer this question, and only 28.9% of farmers said that they had previous experience in credit leverage. This data indicates that farmers are not experienced with credit use. With education and guidance, farmers will become familiar with these credit options, and therefore, they will be more willing to use those instruments to hedge.

The pilot program was exercised by game design, which indicates that given other situations unchanged, farmers will be provided with different products in insurance, such as traditional loan, RCC, and no insurance. The goal is to evaluate the feasibility of RCC and understand the behaviors of potential customers. After understanding the mechanism of RCC and insurance, survey data indicates that the insurance has great potential demand among farmers (Shee, Turvey, Woodward, 2015).

2.4 Other Consideration in Design RCC

2.4.1 Internal Risk Management Method

Small farmers have difficulty applying innovative technologies (such as those in fertilizers and machinery) to improve their production efficiency. To improve the demand for, and efficiency of, the insurance, there are many accompanying interventions.

Small farmers adjust their budget when they are facing extreme weather events, and they need to limit their spending on other expenditures. For example, to maintain an income level, farmers may take actions, such as skipping meals or removing their kids from school, that are obviously harmful to production in a long run.

Even though weather-index insurance is a powerful tool to solve poverty issues in Africa, researchers still need to find other complementary approaches to execute together with insurance. Finally, regulations and laws are necessary in the promotion of this product from pilot experiment to commercial and public level.

2.4.2 External Risk Management Method

Education and professional personnel are needed in the implementation of this product, because farmers have no knowledge of the insurance. For example, if this year doesn't have extreme events, such as drought and flood, according to the design of the insurance, the trigger of the insurance won't be triggered. Small households may think that insurance companies are cheating them, and this misunderstanding will discourage them from buying the insurance in the next year.

Farmers in Africa tend to build a portfolio to diversify risks during their production by growing different crops or having both crops and livestock. Single crop insurance may not be enough to cover all risks exposed the market. Poor transportation may also prevent farmers from getting physical access to credit.

If households understood the mechanism of the insurance by attending classes held by either the insurance companies or the local government, they would be more willing to purchase this product to hedge risks. Education can eliminate confusion and misunderstanding among farmers when the insurance company introduces weather-index insurance. There are many internal factors, such as basis risks, that will also influence the acceptance of the insurance. If the correlation between the indicators, such as rainfall, temperature, and actual environment is not strong the insurance's effectiveness in risk management will be stymied and the insurance won't attain its expected performance. For example, suppose that indicators show that during this crop year, an area has no loss, but farmers do suffer from crop failure. This discrepancy will damage the credibility of the insurance and dampen the acceptance rate of the product.

Spatiotemporal adverse selection is another factor the insurers will have to consider, which will cast considerable influence on final acceptance of the insurance. It is believed that if farmers can have some methods to know the performance of next crop season in advance, they will have a pre-diagnosis about whether to buy the insurance or not (Jensen, Mude, Barrett, 2014).

For example, farmers can use social media or experience to predict the weather of the following season and therefore they can make decisions on whether to use the insurance for next production season. If they have a promising expectation of the next season’s weather, they will decline the insurance. This spatiotemporal adverse selection will be burdensome to insurance company, because if farmers’ prediction is correct with a high probability, insurance company can’t make a profit during good years and only makes payments during bad years, which will make it impossible for the insurance company to generate enough profit to break even in the long run.

Finally, both internal and external factors will influence the acceptance rate of the insurance and the profitability of the insurance company. According to results from the underlying pilot program, current demand for the insurance is low, and the number of farmers who intend to buy the insurance is low.

2.4.3 Basis Risk

2.4.3.1 Local Basis Risk

It is not hard to understand that a pleasant weather condition leads to high production, and local farmers can use weather contingent contracts in mercantile exchange to hedge. However, the imperfect correlation between the contract, which is an indicator of weather, and production, will cause a failure in the hedge. The correlation embedded in the contract is likely to be different from the actual correlation. For example, different crop growth stages have different requirement of rainfall, and so different farmers may be impacted differently by rainfall events.

For example, agricultural insurance in Kenya insures the sum of the rainfall within a specific period, whether during the long rain or short rain. However, the sum may not capture the specific rainfall requirement of crop accurately, since farmers may need more rain in a particular period to keep their crops alive. Even though total rainfall over the year was enough, if the crops die due to insufficient rain in early months, then the farmer is still harmed, but cannot be compensated by insurance.

2.4.3.2 Geographic Basis Risk

The mismatch between the insured areas and the areas which are covered in the contract will also cause basis risk. So, if farmers use a non-local city contract to hedge themselves from weather shock, it won't be an effective hedge for their productions. The payout of nearby location is a good reference to local payout (Norton, Turvey, Osgood). However, due to lack of previous premium, it is impossible for weather-index insurance to apply pricing method.

Small households in Kenya can choose the insurance products that insure an area that has a similar pattern. However, once the insured area and actual production area have a different rainfall pattern, farmers who use this kind of method may have a high chance of suffering a loss without compensation.

2.4.3.3 Product Basis Risk

Farmers can choose different indicators to use in contracts to protect themselves, and different indicators, such as precipitation and temperature, will have different effectiveness in hedging. For example, in cotton production, temperature may play a more significant role than precipitation. However, the quantitative correlations are hard to measure and therefore, it can be difficult to find a concrete number for each contract. For cotton producers, if rainfall and temperature are weighted equally in the contract, it will cause a failure in hedging.

After understanding pattern of a single indicator, more indicators and correlations would also be considered in model construction. For example, rainfall may not be the only factor that affects crop production. Temperature and other indicators also have an influence on production, but more indicators and correlations make it more complicated for researchers when designing the product. The introduction of new and effective indicators will help to capture typical rainfall pattern of local area.

2.5 Related Program

Before weather-index insurance, there are other commodity-linked credit instruments in Kenya. Big farmers, manufacturers, and farmers who have strong capital foundation and required knowledge, will use commodity-linked bonds, which increases their exposures to capital markets, and reduces their financial risk (Turvey, 2006).

2.5.1 Direct Cash Payment Method

HSNP (Hunger Safety Net Programme) is more efficient to the poorest farmers, because they are the group of people who even don't have enough capital to purchase the insurance and don't have the productive tools to maintain a sufficient daily production. Insurance and subsidy won't benefit them directly. Ultimately, these improvements will lead to an overall increase in agricultural productivity. In a long run, improvements in agricultural sectors will lead to improvements in non-agricultural sectors, too.

The cost of these programs is a critical issue that researchers cannot avoid considering. Operating and administrative costs, such as expense in marketing and monitoring, are high. The duration of direct cash payment is uncertain, because a cash payment program only wants to help the poorest farmers to reach a sustainable level of production. Certainly, the program needs to last for a long time and have a strong funding foundation so that it can support farmers to become independent gradually and finally get out of poverty trap. The capital requirement can be burdensome for local government and donor communities.

Opportunity cost is another factor that needs consideration when apply direct cash payment method. That's to say, the funds used for cash payment may help the poor more effectively if applied to other more promising and profitable programs. This "misdistribution" in resource arrangement will affect the development of the whole economy.

An insurance program applies different methods from a direct cash payment program, and it will be discussed in detail in a later part. Insurance can reach comparable results to direct payment when farmers are facing a catastrophic weather event, and the households who benefit from this product are generally different from the group of people who receive the benefit from direct cash payment. The selection of these two methods will be determined by clients' situations and budget of the program.

Currently, political intervention plays a significant role in dealing with covariate risk when farmers are facing extreme weather events. For example, in Peru, there is a debt forgiveness policy, which significantly increases the default probabilities of many borrowers. This intervention causes an unhealthy feedback loop between the unwillingness to issue similar instrument from insurance issuers and higher and higher default rate from customers.

Sub-Saharan Africa is a place that is vulnerable to weather fluctuations, because there are many small and poor households living there, and their income comes predominantly from agricultural production, which is very sensitive to weather shocks. The high frequency of climate change and increasing occurrence of extreme weather events makes farmers more exposed to risks than before.

Technically speaking, in agriculture, there are two methods to cover loss caused by these unpredictable weather patterns, which are risk minimization and loss management. Risk minimization refers to the plans and strategic decisions made before production, such as crop diversification and intercropping. Loss management refers to the makeup reactions after certain shocks, such as off-farm employment and self-insuring behavior. Weather index insurance can be considered as a risk minimization method, because it is the decision that farmers are going to make before the following crop season with the intent to reduce potential losses in future.

2.5.2 IBLI (Index-Based Livestock Insurance)

IBLI program focuses on assets, which are livestock, and this mechanism provides a reliable protection for farmers to maintain a sustainable level of production. Assets are necessary to generate income. If fundamental assets are damaged severely, there is low possibility to have a stable cash flow now or even in the future. Especially in agricultural production, loss of productive assets will put or lock small households into the poverty trap. According farmers' experience and previous research, livestock's mortality is affected significantly by external environment. Farmers who bought IBLI will be paid according to the difference between benchmark and their actual livestock mortality. This is a dynamic process, and farmers may get different amount of repayment based on different weather conditions during the insured time interval.

When designing the index-based insurance, researchers and local governments consider scalability and sustainability of this product as well. Scalability refers to the expansion of the product from small pilot project to a standardized commercial product and a broader market. Sustainability refers to that RCC products should have a long-term viability in commercial markets, which requires that it should both benefit customers and issuers. If the premium of the insurance is too low, issuers may have trouble breaking even after paying the claims on the insurance. If the premium goes beyond the customers' maximum willingness to pay, the demand for the insurance will be too low.

The mechanism of the insurance, social infrastructure construction, and regulatory enforcement will influence the sustainability of the insurance. For a sustainable commercial product, its premium should cover all costs of providing the insurance. Premium can be divided into two parts. One is pure premium, which refers to the expected payments to total amount of money insured. This is the cost that farmers should pay in advance to protect themselves and get paid in the future, if a drought or a flood happens. Another part of the premium is the operating cost, which refers to the expenses incurred when issuers design and administer this product, such as marketing, training, and data collection.

In Africa, livestock is a major income resource for households, and it can be significantly affected by weather, such as rainfall and temperature. According to previous research results, drought-related factors accounted for 53% of the livestock deaths. Disease, which is correlated with drought will, accounted for another 30% of deaths during the same period (McPeak, Little & Doss 2012). Therefore, it is easy to see the importance of index insurance, because livestock's mortality rate is positively correlated with weather indicators, such as rainfall and temperature.

The methodology of this insurance is transferable. However, weather-index insurance is still hard to replicate because different areas will have different prominent characteristics that influence the local environment. Ideally, researchers will want to identify the most relevant ones when setting the benchmark.

According to the results from IBLI program (Jensen, Mude, Barrett, 2014), due to macroeconomic uncertainty and cultural preferences, farmers tend to keep livestock as a store of value. However, livestock is an asset with low liquidity, and so this method has an obvious shortage. According to the laws of supply and demand, if quantity supplied, given the price, is higher than quantity demanded, given the price, then there is a surplus of the product, and the price of product will go down. Similarly, for livestock market, if covariate shocks happen, all nearby farmers will tend to sell their livestock to liquidate the asset and therefore, the price that farmers can get by selling livestock now will be lower than the price that they would get in years without weather shock. Therefore, this method is not be an effective way to maintain a stable cash flow. Due to the instability of self-insured method, some farmers may prefer to take advantage of external methods to hedge risks, like weather-index insurance.

IBLI is a regular pilot program in Kenya, which is sold twice each year. There are two windows for farmers to purchase the insurance. These two purchasing times are the long rain season, which is from January to February, and the short rain season, which is from August to September. Therefore, farmers will have flexibility to choose which time periods they want the insurance to cover.

During the pilot program of IBLI, some problems occurred, which will also need attention when we are dealing with the design of weather-index insurance.

The premium charged to small households is a critical issue that researchers must bear in mind when they are designing with this financial instrument. According to the IBLI pilot experiment, farmers are sensitive to insurance premium changes, which indicates that a lower premium will attract more potential customers. The elasticity of this instrument is relatively high. Compared to a high potential repayment of the insurance, farmers are more interested in an insurance product with a lower premium, even if the potential repayment is lower.

High premiums charged by insurers may still be attractive to highly productive farmers; however, they are not the targeted customers who will achieve the most gain from this instrument, because the target of this product is small households who are struggling in the poverty trap, while wealthy farmers are far beyond the threshold of poverty trap. Therefore, the premium charged by this type of insurance should be attractive to its target customers, and different premium level will attract different groups of farmers who belong to different income levels.

Looking at the results of this program, there are some findings that can also be applied to RCC. It was designed to lessen damages caused by extreme events. In the long run, it was designed to help farmers get rid out of the poverty trap (Barrett, Barnett, Carter, 2007).

According to pilot result, the demand for the insurance is low. Due to the lack of the awareness of the insurance and its characteristics, small farmers tend to undervalue the product, and they are unwilling to accept an expensive but correctly priced product. Subsidy is a problem that researchers can't ignore during their work. The original intention of the subsidy is to lower the premium to increase demand. After farmers are getting experienced with the insurance and issuers gain more practical experience, subsidy can be eliminated so it becomes a competitive commercial financial instrument, which can be openly traded on the open market. The impact on the valuation due to a subsidy should be examined, and if this assessment process can't be evaluated properly, then the market will tend to have negative reactions towards these subsidies. The introduction of financial intervention should be careful and cautious. Otherwise, it will jeopardize the long term sustainability of the product.

2.6 Lessons

However, there are many traps that need attention. The premium can't reflect the actual value of the product, and there is a high chance of making economically inefficient decisions. When donors remove the subsidy, insurance companies may not be very confident about the number of loyal customers who are going to stick to the insurance even though the price of the insurance is higher than before. In addition, the introduction of this new product may dampen use of other risk transfer instruments on the market. If this problem can be addressed properly, other components in risk transfer instruments can combine to cover the area where index-insurance fails to be appealing.

For insurance companies, who want to earn profit from these instruments, they may have concerns when they enter this new market, due to low price tolerance from local farmers. This premium should go below the maximum willingness to pay from their customers. If the price charged by insurance companies is above the willingness to pay of customers, a subsidy will have to be introduced. Subsidy is a common solution to limited capital, because it is an efficient way to stimulate both supply and demand. With the help of the subsidy, insurance companies can charge a higher price than the price without subsidy, and small farmers will be offered a comparatively lower price than the one without a subsidy. Subsidy will be responsible for the difference between these two parties. Subsidy providers, such as government and international institutions, can be a reliable backbone for insurance companies, because when an extreme weather event strikes, they can use ample capital and resources from themselves to protect insurance companies from collapse caused by covariate risks. However, subsidy may bring potential damage to the market. It may crowd out existing products due to price competition. It may turn out to be that insurance products without subsidy are all eliminated from the market, due to their higher price to consumers. In a long-run, subsidy distorts the operation of an efficient market, and when subsidy is removed from the market, existing financial products may have difficulty supporting themselves. The total number of insured customers in a specific location is an issue that needs consideration when government or donors subsidize the index-insurance.

The burden of the insurance might be heavy for insurance companies, which is also the reason why insurance needs subsidy. There is another way to ameliorate the burden put on insurance company, which is layered-payment, that’s to say, different parties will be responsible for different parts of the insurance repayment.

Small amount of loss within some budget can be covered by insurance companies themselves; however, when the loss goes beyond the expected budget, which makes the payment unaffordable to those insurance companies, governments can take responsibility to pay the rest part of repayment. Risk-layering can help insurance company transfer risks when their portfolios are not widespread enough to diversify systematic risks. Government or NGO will become their last line of defense, and this mechanism will provide an opportunity to share the loss and burden among different parties. Consequently, no specific party will suffer a devastating loss, and players who get involved to this index-insurance all can maintain a sustainable business.

Obviously, farmers can take advantage of this financial instrument to equip themselves with more innovative and efficient financial instruments to maintain a stable cash flow in the future. Compared to direct cash payment, which is another popular way in Africa for the government to relieve poverty, insurance can apply indicators, which are easy to measure and highly related to agricultural production, to calculate repayment flexibly, because the repayment will be adjusted to actual environment, which is more flexible than direct cash payment method. Index-based insurance will decrease the moral hazard and adverse selection problem effectively. It doesn't require client's historical credit record, which will be helpful to insurance companies, because small households, especially those who live in extremely impoverished areas, don't have these reports for insurance companies to refer to.

Another interesting finding in the IBLI program is that there are different discounts provided by issuers to reduce households' final premium to motivate them to purchase the insurance. For example, during each sales window, a randomly selected coupon which ranges from 10% to 60% will be applied to farmers' premium to reduce farmers' cost when they are purchasing the insurance. This is another way to increase the demand for the product. In the initiation of RCC product, similar strategies can be applied here to motivate the demands of those innovative RCC-embedded financial instrument.

With the application of RCC, the risk transfer method can help farmers become more financially flexible for farmers because with the help of the insurance, farmers don't have to rely on comparatively costly self-insurance method. The initial investment required to provide the insurance is huge, such as the construction of weather stations and high quality data collection infrastructure. There is a huge amount of preparation required before the application of the index insurance, such as education of customers and marketing to increase awareness. If the customer base of the insurance isn't large enough, insurance companies will face great cost per capita. However, to guarantee the fitness of the insurance, insurance companies can't apply same models to physically widespread areas. If these areas are too far away from one another, environmental conditions can vary widely. The tradeoff between greater diversification by spreading to new areas and the cost of estimating new models is something researchers must face when they are trying to implement this financial risk transfer instrument.

For the utility industry, structured financial products also has its own benefit. These industries are sensitive to weather, and structured financial products provide a feasible way to eliminate or decrease price fluctuation risks. Meanwhile, this strategy also solves liquidity problems and mitigates downside weather-related risks. RCC applies the similar strategy here to minimize the volatilities during the agricultural production. Farmers will always maintain a stable cash flow either during the harvest season or drought season. During the harvest season, farmers can support following season's production by selling crops, and during the drought season, farmers can get compensated according to the loss caused by drought from insurance companies by purchasing RCC-embedded production in advance.

The goal of this dynamic method is to protect farmers from the poverty trap by providing them a more accessible credit and capital market. Weather risks put small farmers into the poverty trap, and limited access to capital market makes it difficult for farmers to get out of the trap. Under current situation, RCC can be an effective method to minimize the loss caused by downside risk, which is rainfall here. A complete and efficient access to the credit and insurance market can benefit both small farmers and issuers. If products can be sold through existing channels, such as existing insurance companies and informal insurance groups, it can be more widespread and more easily learned about by potential users. Farmers can use this access to know more about the insurance, so that they are more willing to invest in weather-index insurance. Correspondingly, for insurance issuers, they can also use this reliable channel to provide products and other accompanying service, such as education and product usage training For small farmers who live in remote areas, the cost of traditional claim-based insurance can be high. Insurance companies at this time can make better use of mobile devices to reduce transaction costs and increase access to the product and service. In addition, a successful and complete risk management system also requires the enforcement of related laws and regulations, which is commonly nonexistent in those countries.

Chapter 3: Data source, Characteristics and Time Series Method

3.1 Introduction

In chapter 2, I provided an overview of the conditions which are currently being faced by farmers in Kenya and proposed the structure of an RCC product to be applied in a pilot program. The RCC trigger is based on cumulative rainfall, using the 20% quantile of pert distribution.

The purpose of this chapter is to provide an alternative method to RCC using time series model (Robert, David, 2010). A time series is a set of data recorded in time order. It is discrete and equally spaced. Time series forecasting is the use of past observed values to predict future values. Models for time series tend to have a better performance in a shorter term than a longer term. Therefore, a model based on recently observed data will be more accurate than one based on data from long time ago. Therefore, forecasting will be more accurate. Therefore, this paper uses more recent data to improve the accuracy of the forecast.

Machakos County is a semi-arid and hilly terrain areas in Eastern province of Kenya. The rainfall of this area is very low, which is around 700 mm per year. (Situation Analysis-GOK 2014). The average rainfall in long rain and short rain is 315 mm and 266 mm respectively. Small households in this areas are unable to do self-hedging due to their low wealthy level. Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS) is a quasi-global rainfall dataset, which provides the historical rainfall data of Machakos County in past 30 years.

The goal of this chapter is to construct a seasonal time series model to reflect general pattern of rainfall in the pilot area. Insurance companies and local microfinance institutions can refer to this benchmark to price the insurance.

Kenya lies in east of Africa. There are two rain seasons which compose a major percentage of rainfall in Kenya, according to local farmers' experience. The period from March 15th to May 15th is the short rain, and the period from Oct 15th to Jan 15th (of the next year) is the long rain.

Figure 3.1: Monthly Average Rainfall Results Across 35 Years

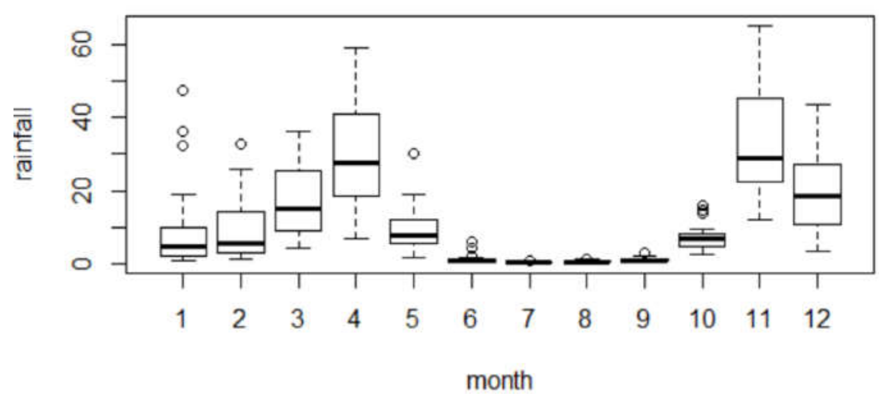
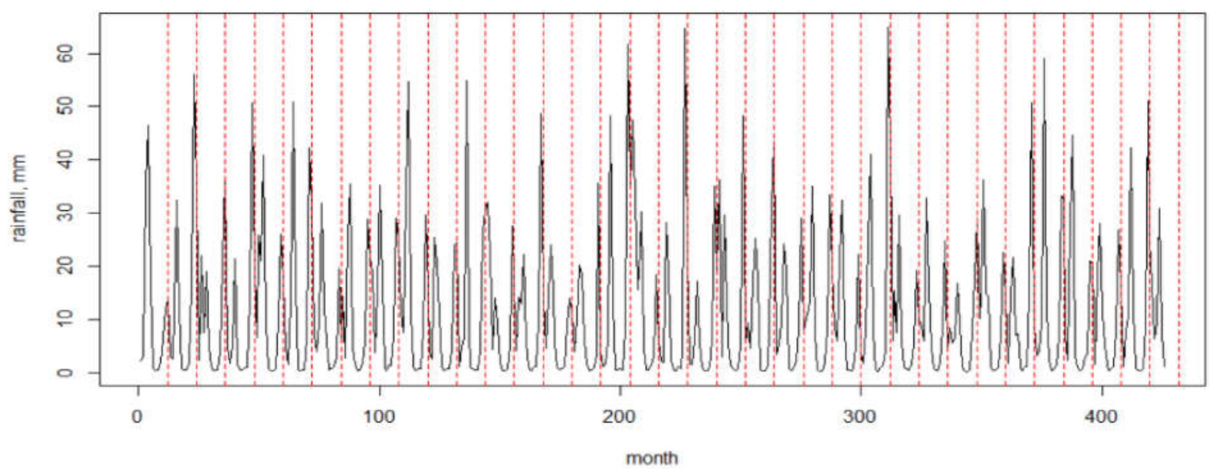


Figure 3.1 shows a cyclical and seasonal pattern of sample plots in a year. This pattern confirms that the long rain season and the short rain season compose a major percentage of rainfall within a year. In fact, the long rain season and the short rain season account for a majority of the rainfall, and variance during these two major rain seasons is large. Therefore, this adds uncertainty to farmers' agricultural production. For small households, who lack capital to apply a sufficient self-insured method, will seek external support to protect themselves from extreme weather events.

The data for this model is biweekly observations of rainfall from the past 35 years. The total number of observations for this area is 420. The total rainfall for a given month is simply the sum of the individual rainfalls for that month.

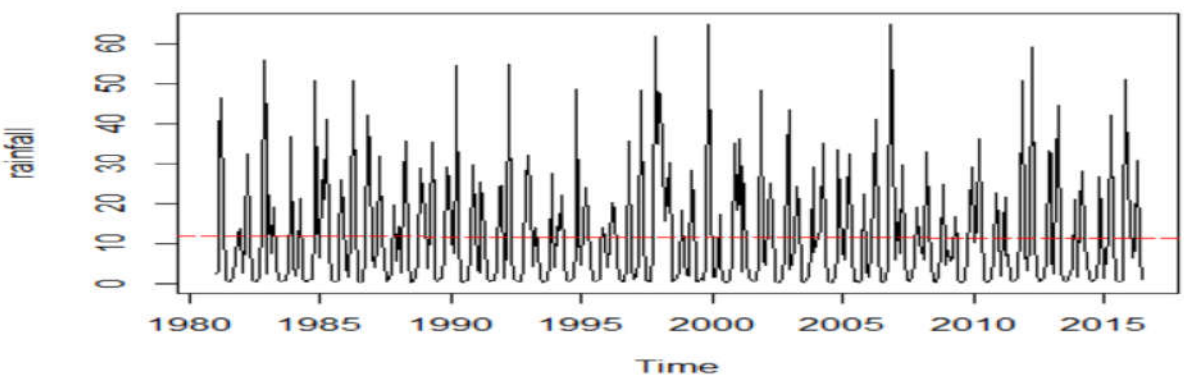
To capture the monthly pattern of the rainfall data, I reconstructed the data into 12 points for each year by taking the average of the rainfall within each month. The rearranged data are equally spaced, and therefore can apply corresponding time series model. After understanding patterns of each month, results from short rain and long rain will be grouped up to create trigger to exercise this insurance. In addition, differencing $(X_t - X_{t-1})$and log transformation may be applied here to improve the accuracy of the model.

Figure 3.2: Time Series Plot of Monthly Rainfall for Example Area Across 1981 to 2015



In Figure 3.2, the interval between two dashed lines indicates the length of year. Within each period, we can tell that there is a peak and trough. Even though the exact magnitude of rainfall is different from year to year, we can still detect annual patterns from year to year. For example, during the cycle, there is significant high rainfall and low rainfall, which provides the possibility in which we can use past rainfall pattern to forecast future rainfall pattern.

Figure 3.3: Time Series Plot of Monthly Rainfall for Example Area Across 1981 to 2015 with Trend Line



Another regression line is drawn in Figure 3.3 to check general trend of rainfall. The Figure 3.3 shows no significant increasing or decreasing trend in total rainfall. Constant mean is one of three pre-conditions of stationarity. If there is a trend, either increasing or decreasing within data, we should apply extra methods to remove this trend.

3.2 Model in Pilot: Pert Distribution

In the 2017 and 2018 pilot program, PERT distribution was applied to simulate rainfall in Machakos County. In the simulation model, a lower 20% quantile of the rainfall was selected as the local trigger. Time series method applied here provides another way to detect local trigger. After getting results provided by two methods, we want to compare those two sets of results to see the difference or the similarity.

The rainfall data of Kenya in 11 plots are provided, from 1981 to now. In each rainfall season, cumulative rainfall and average rainfall are recorded. Even though these two sets have different numbers, they have the same shape since the only difference in these sets is the scale. These numbers are going to be incorporated into the rainfall simulation later.

3.3 Correlation and Covariate Risk

One of the most critical elements of RCC is the correlations among plots in Machakos County. If correlations among plots are strong, it indicates that the drought of one plot is a strong indicator of the droughts in other plots.

Within the long rain season, rainfall levels are highly correlated. This implies that if drought occurs in one plot, other plots are highly likely to have drought, too. This results in a potentially high covariate risk. Drought in one plot can have a significant influence on other plots. The Insurance Company should therefore pay close attention to its portfolio diversification.

A similar matrix can be constructed between the long rain season and the short rain season, and, between short rain seasons. The correlations within the long rain and the short rain are strong across plots. The Figure 3.4 and Figure 3.5 show correlation within long rain and short rains, respectively.

Figure 3.4: Long Rain Correlation Matrix

Figure 3.5: Short Rain Correlation Matrix

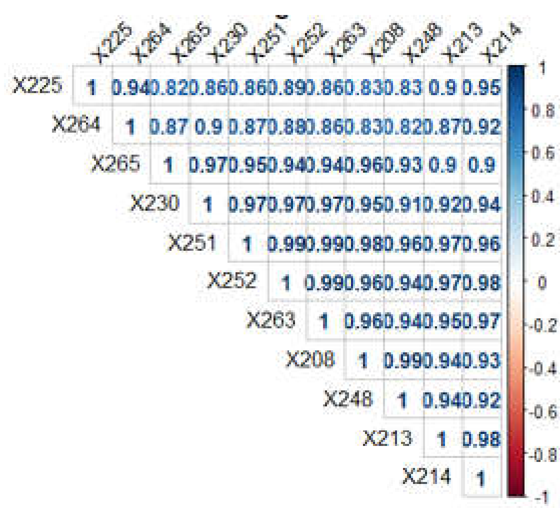
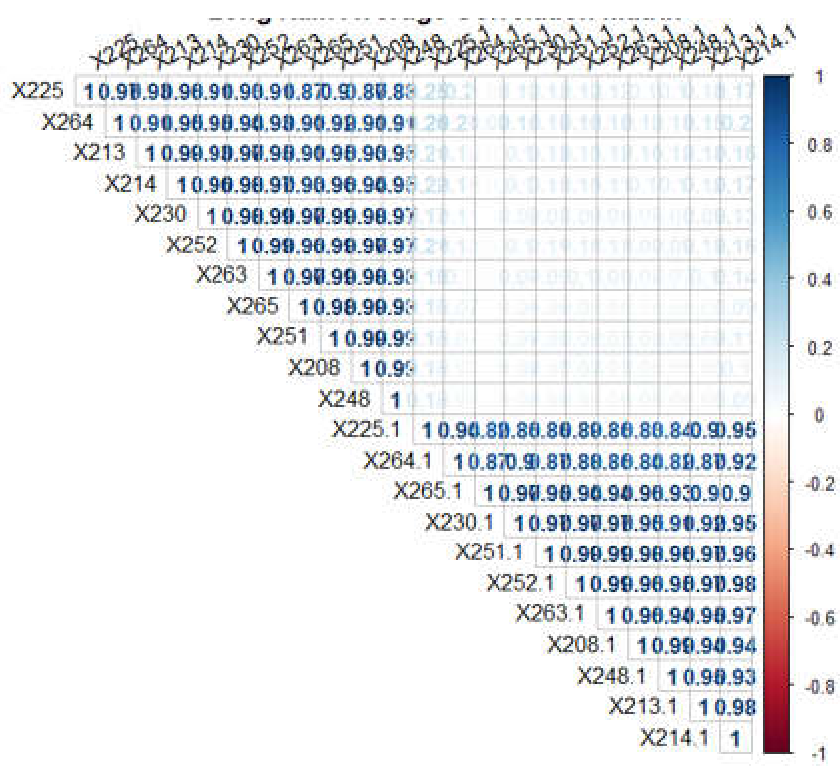


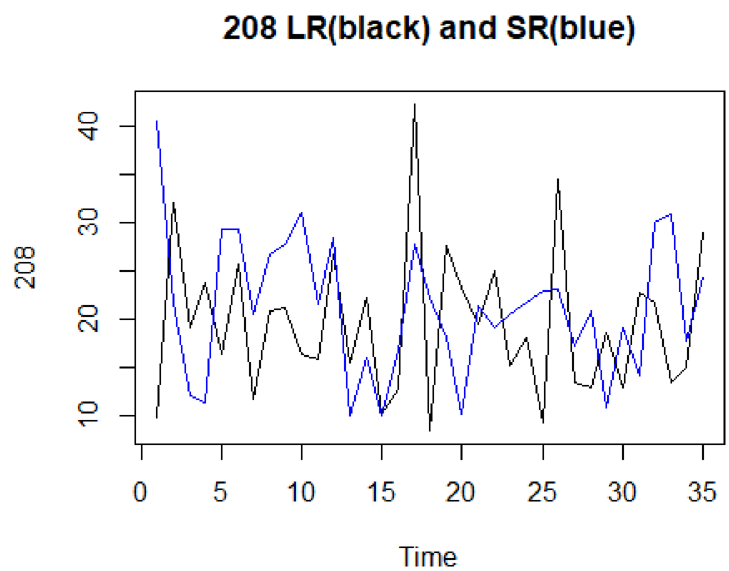
Figure 3.6 shows the correlation of long rains and short rains:

Figure 3.6: Long Rain and Short Rain Combination Correlation Matrix



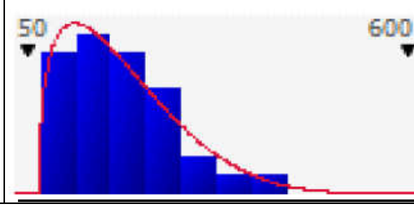
According to Figure 3.4, Figure 3.5 and Figure 3.6, we can tell that the correlation within the short rain and the long rain is strong. Correlation between long rain and short rain is positive, but it is significantly weaker than the correlation within a season, which indicates that a drought during the short rain won't be a strong predictor of drought the following long rain, and vice versa.

Figure 3.7: Long Rain & Short Rain Overlapping, Central Machakos



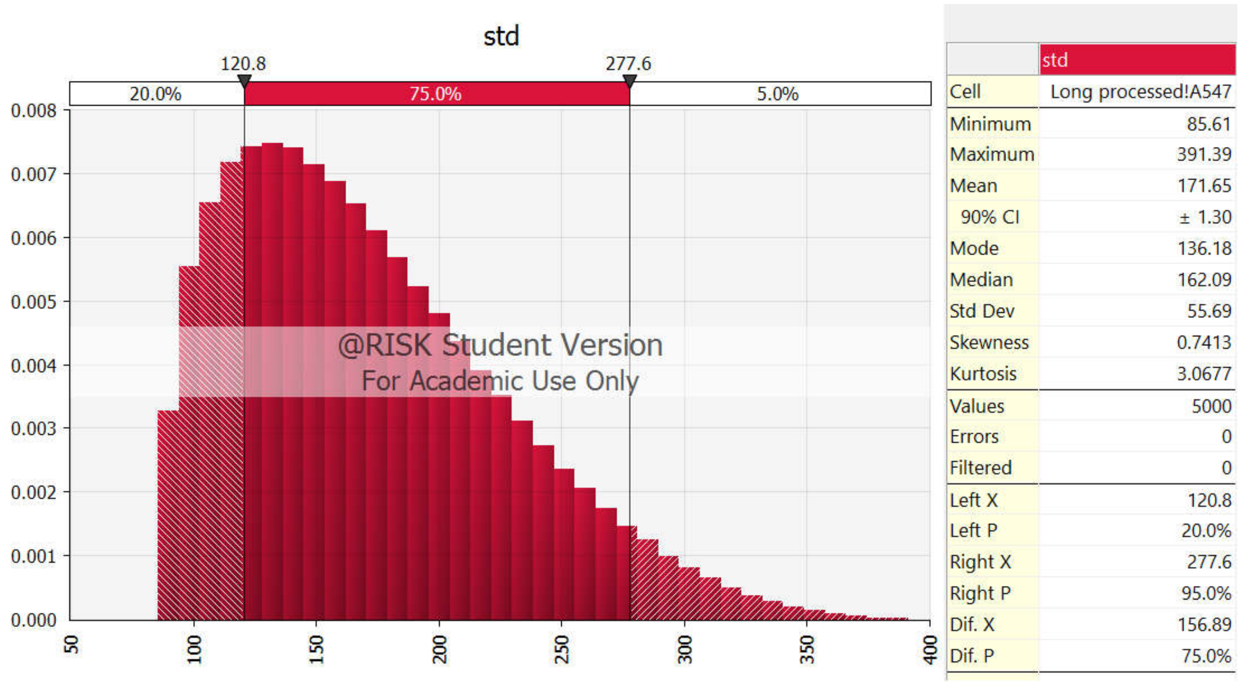
In the construction of the models, the pert distribution is chosen to simulate rainfall. This model has three parameters, and they are min value, mode value, and max value. Max and min rainfall can be found in the historical cumulative rainfall data. After defining the shape of model in @Risk, a mode for the rainfall will be generated automatically. Finally, these three parameters will be used in the following rainfall simulation and trigger determination.

Table 3.1: LR Plot 208 (Central Machakos): Simulation

Min	85.306994
Max	423.047899
Mean	195.0542
Mode	130.3793
Median	180.1818
Std. Deviation	77.1972
Graph	

After processing the rainfall data, the corresponding simulation models are created to simulate the pattern of the rainfall. For each plot, there are 2 models, one for long rain and one for short rain. According to figure below, the lower 20% bound is 120.8, which is the lower 20% quantile of the rainfall. This is the trigger that we are looking for in the insurance premium design.

Figure 3.8: Pert Distribution Simulation



3.4 Data Transformation with ARIMA

Obviously, observations in a time series are rarely independent. We should test the series to verify that it is stationary. After series passes this test, we can then apply the time series method.

ARIMA (Autoregressive Integrated Moving Average) methods are widely used in time-series analysis and forecast. It has three components: autoregression (AR), differencing (I), and moving average (MA). Before applying ARIMA model, we should check stationarity of the data set. If data is not stationary, we should increase the level of differencing. If we ignore the trends within the data, this will damage the accuracy of the model. For example, the sample mean generated from previous data will fail to capture known patterns when predicting future values, if we don't take trend into consideration in the model.

Especially, for seasonal ARIMA model, it final presentation should be: ARIMA(p,d,q) (P,D,Q), where p is parameter for the non-seasonal AR model, d is differencing times for non-seasonal, and q is parameter for the moving average component. Capital P, D, and Q indicate that parameters for seasonal patterns. Seasonality is a regular pattern that repeats over a fixed time periods.

Given that there is no differencing operation, a stationary process Y_t , SARMA(p,q)(P,Q), can be written as:

$$(I - \phi_1 L - \dots - \phi_p L^p)(I - \Phi_1 L^s - \dots - \Phi_P L^{sP}) Y_t$$
$$= c + (I - \theta_1 L - \dots - \theta_q L^q)(I - \Theta_1 L^s - \dots - \Theta_Q L^{sQ}) u_t \quad (1)$$

Where u_t is a white noise process.

In the selection of parameters, we should first find s, the order of the seasonality. We may first use autocorrelation function and partial autocorrelation function to find an MA(q), or AR(p) in non-seasonal part.

If there are significant correlation of the s order, then we can select parameters for the seasonal part of the function. For example, we could use either SARMA(0,q)(0,1) or SARMA(p,0)(1,0). The difference between SARIMA model and SARMA model is differencing. During the data transformation process, if there is seasonal differencing between data, we should use SARIMA model. "I" in the model stands for "integrated". If there is no seasonal differencing, SARIMA model and SARMA model should be equivalent. If there are still significant correlations around 2s, then we should consider a SARMA(0,q)(0,2) or SARMA(p,0)(2,0). However, if none of models above work, we will begin to combine AR and MA terms. For example, SARMA(1,0)(0,1), SARMA(1,1)(1,0) and so on.

3.4.1 Stationarity

There are three conditions that you should check for stationarity.

1. Constant mean for all time periods. If there is significant trend, you should use differencing techniques to remove the trend.
2. Constant variance for all time periods. If variance is increasing or decreasing with time, you should not consider the variance as constant.
3. Autocovariance function between period t and period t+k should only depend on the interval, k. Any other interval with the same length should have the same autocovariance. This function tries to capture the dependency structure of the process.

- Autocovariance

$$\gamma(s,t) = Cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)],$$

for all $s, t \in \mathbb{Z}$.
For $k = 0$

$$\gamma(t,t) = Cov(X_t, X_t) = Var(X_t). \tag{2}$$

The autocorrelation of order k can be defined as following equation:

$$\rho_k = \mathrm{Corr}(Y_t, Y_{t-k}) = \frac{\mathrm{Cov}(Y_t, Y_{t-k})}{\mathrm{Var}[Y_t]} \tag{3}$$

If the correlation is high, we can tell that observations within a series are highly correlated, and vice versa. If the time series has a trend within it, it would be necessary to use difference operators.

The difference operator Δ is defined as

$$\Delta Y_t = (I - L)Y_t = Y_t - Y_{t-1} \quad (4)$$

If a stationary process is obtained by applying differencing operation, we say that Y_t is integrated of order 1. For series with a seasonal pattern, we can also apply differencing operation of order S. For example:

$$\Delta_s Y_t = (I - L^s)Y_t = Y_t - Y_{t-s} \quad (5)$$

If s=4, this indicates a quarterly data.

If s=12, this indicates a monthly data.

3.4.2 Procedure

Stationarity of the dataset is the first condition that is checked. After confirming the stationarity of series, next step is to determine the parameters of model. In this step, the parameters can be determined from the autocorrelation function (ACF) and partial autocorrelation function (PACF). The data set will be split into a training set and a test set to do a cross validation to test the accuracy of the model. Meanwhile, the residual of models will also be tested here. It should follow a normal distribution. If there is some other trend, either linear or non-linear, within the residuals, this indicates that model fails to capture some characteristics within the original series. Other data transformations should be applied to capture those previously undetected pattern. R was used to do this time-series analysis and code will be provided in appendix.

There might be several candidate models to fit the series, and after deciding the parameters of the function, a cross-validation will be applied to find the precise parameters of the model to minimize the total sum of squared error. In addition, the Akaike information criterion (AIC) can be another useful tool to help determine appropriate parameters.

3.4.3 Overfitting

A good model should also avoid overfitting problem, and that's to say, the order of this time series model should be as few as possible. This is a tradeoff between the bias and variance. Overfitting can be considered as modelling error, because it is overfitted for the training data, and therefore, it may not fit test set very well. Therefore, the model can become overly complex, and the predictive power of the model will be reduced due to overfitting. Cross validation can help to solve this problem.

3.4.4 Unit Root

The prerequisite condition to apply time series analysis is stationarity, so we must check the stability of time series data. The augmented Dickey-Fuller test will be used here to test this condition. We want to apply unit root test:

$$H_0 : \phi = 1 \text{ versus } H_a : \phi < 1$$

$$Y_t = a + \phi Y_{t-1} + u_t$$

Rejecting $H_0 \rightarrow$ stationarity.
Not Rejecting $H_0 \rightarrow$ Unit root

A unit root indicates a random walk pattern with a stochastic trend, which indicates that it is inappropriate to use time series analysis here.

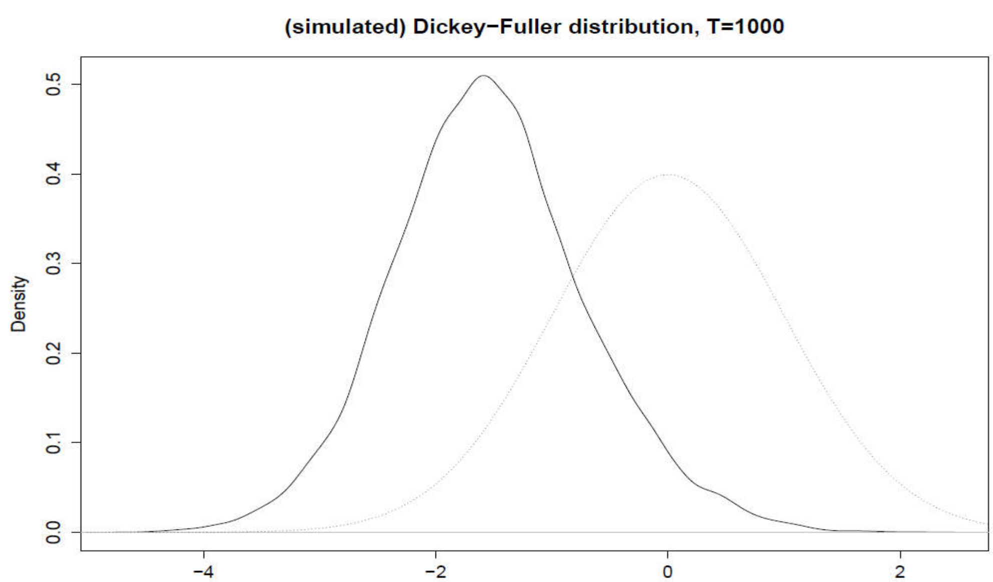
$$\Delta Y_t = a + \gamma Y_{t-1} + u_t \cdot \gamma = \phi - 1 \quad (6)$$

The test statistic is:

$$t = \hat{\gamma} / SE(\hat{\gamma}) \quad (7)$$

H_0: Y_{t-1} is nonstationary. The t-statistic doesn't follow a t, but a Dickey-Fuller (DF), distribution.

Figure 3.9: Dickey-Fuller Distribution



According to the distribution, we can see that the critical values of the DF distribution are more negative than those of a normal distribution. Especially, x here stands for a numeric vector or time series. In a Dickey-Fuller (DF) test, H_0 assumes ΔY_t is a white noise. In Augmented Dickey-Fuller (ADF) test, H_0 is enlarged, and it allows ΔY_t to be an AR(p). Therefore, it can be written as:

$$\Delta Y_t = \mu + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \ldots + \delta_p \Delta Y_{t-p} + u_t \quad (8)$$

u_t : White noise.

So the test will be:

$$H_0 : \gamma = 0 \text{ versus } H_A : \gamma < 0 \quad (9)$$

$$t = \hat{\gamma}/SE(\hat{\gamma}) \quad (10)$$

We still use Dickey-Fuller distribution instead of a t distribution or a normal distribution. When the Augmented Dickey-Fuller test (ADF test) is applied, the p-value of the test is 0.01, which is smaller than 0.05. The null hypothesis of the test is that the series has unit root; alternative hypothesis is that the series is stationary. Based on the result of the test, we consider the data set is stationary. There is no need to use differencing, but the log transformation is taken to minimize the variance of the series and improve the accuracy of the model.

3.4.5 Rainfall Differencing Operation

An important pre-requirement of using time series method is the stability of the series. One method is to see the distribution of year of year differencing distribution. The null hypothesis of the test is that the distribution has a mean 0. This indicates that the year of year differencing is zero, and therefore, the series is stationary. There is no significant seasonal trend within the series.

The graph below is the distribution of seasonal differences in rainfall between t and t-1. For example, it can be the difference between January 2013 and January 2014 at plot 208 (Central Machakos). The distribution has a mode of 0, and there is no significant increasing or decreasing trend within the distribution.

Figure 3.10: Seasonal Differencing Distribution, Central Machakos

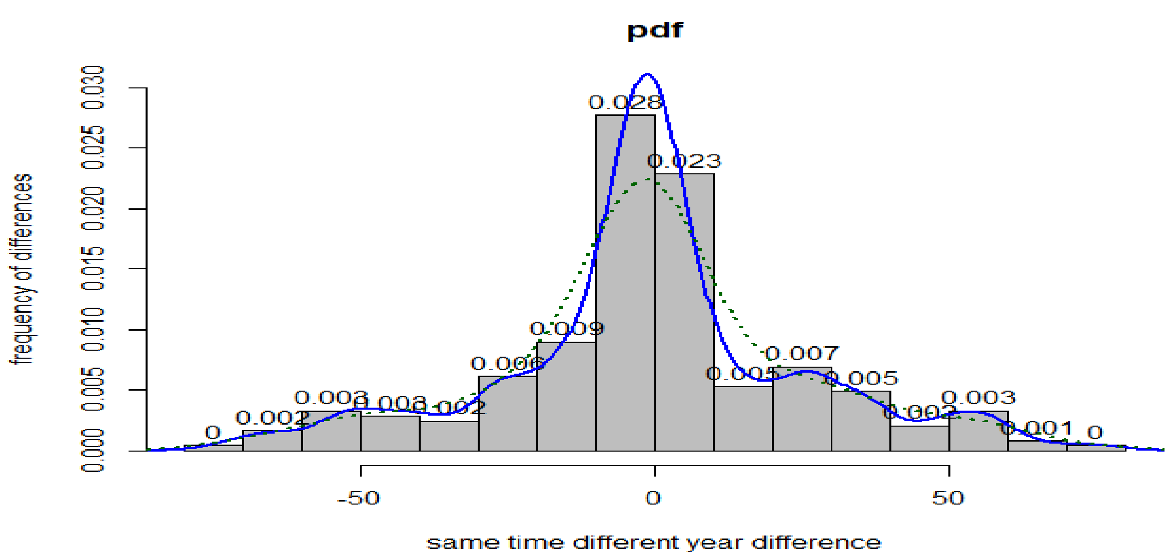
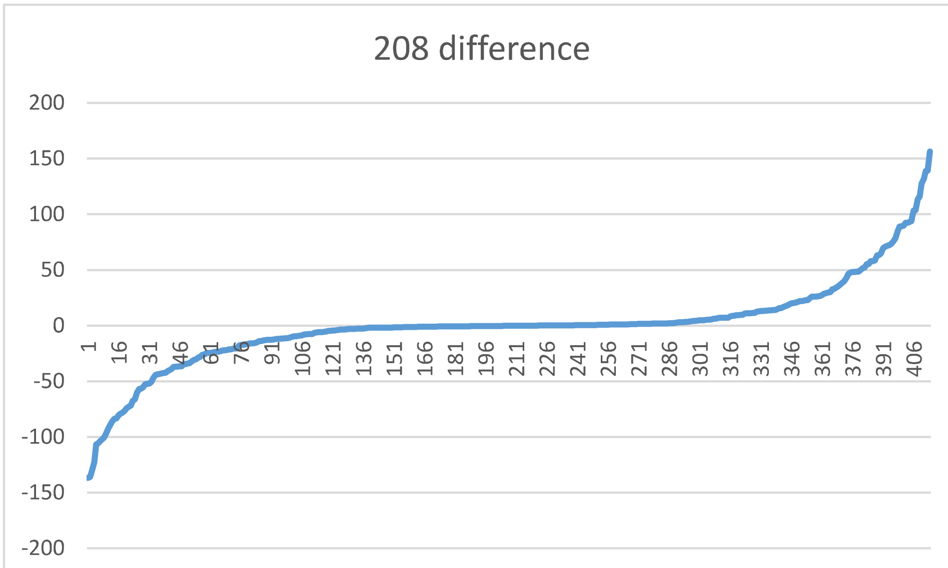


Figure 3.11: Cumulative Distribution of Seasonal Differencing, Central Machakos



After calculating seasonal differences, and rearranging the data from the smallest to the biggest, a line is drawn according to this. A high percentage of data falls around 0. The highest difference is around 150mm.

Seasonal differencing graphs of other areas are listed below.

Figure 3.12: Cumulative Distribution of Seasonal Differencing, Yathui

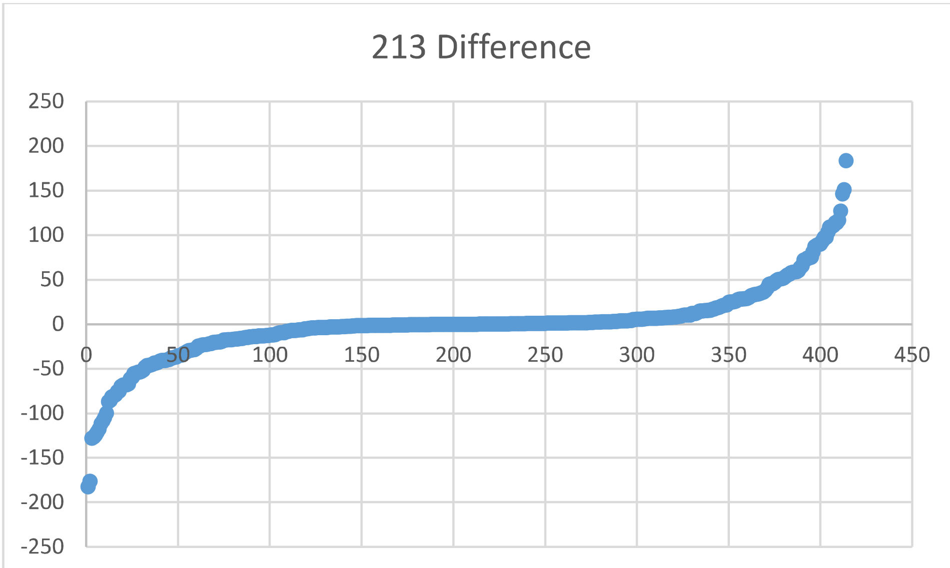


Figure 3.13: Seasonal Differencing Distribution, Yathui

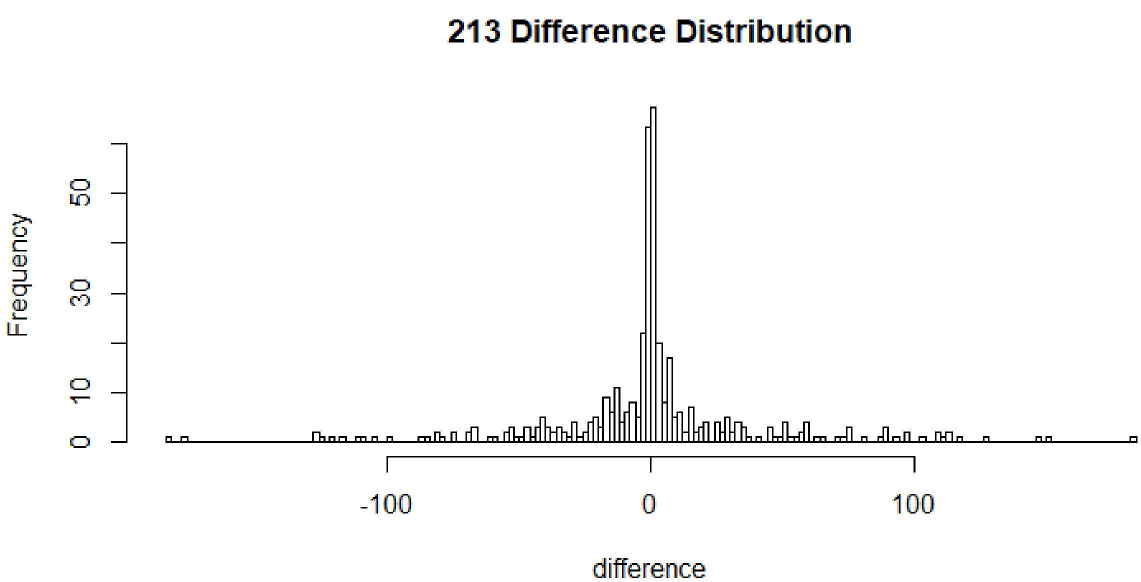


Figure 3.14: Cumulative Distribution of Seasonal Differencing, Yatta

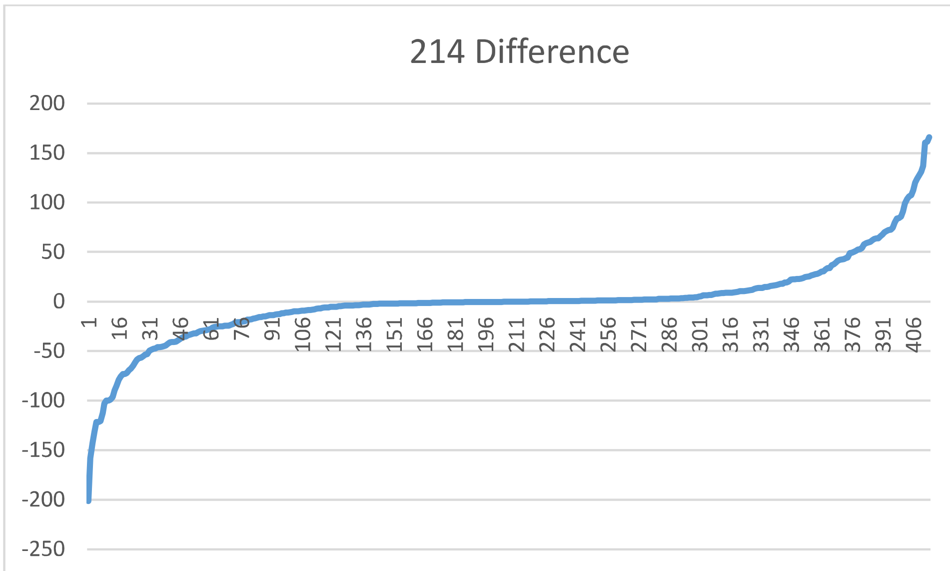


Figure 3.15: Seasonal Differencing Distribution, Yatta

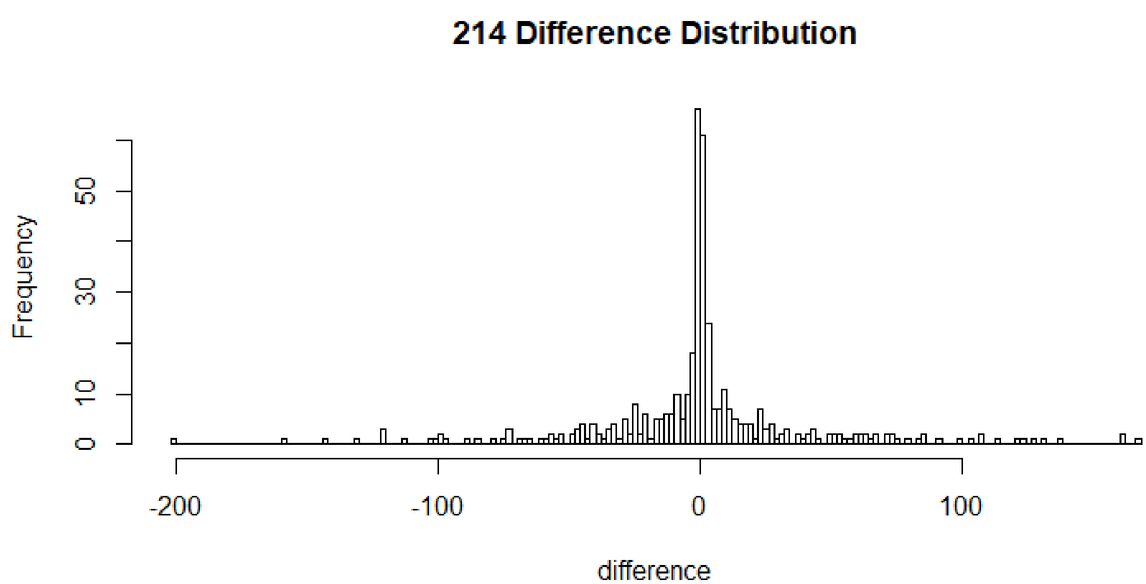


Figure 3.16: Cumulative Distribution of Seasonal Differencing, Masinga

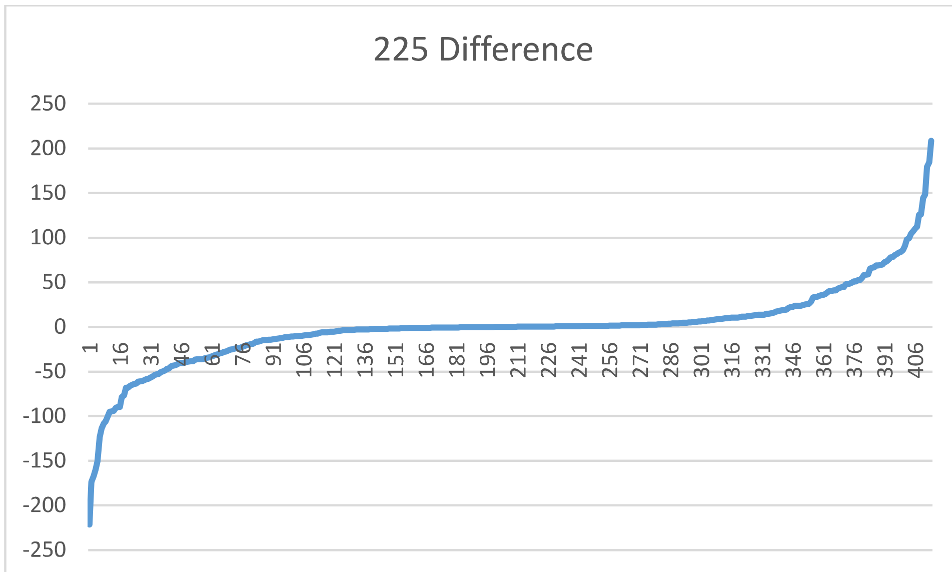


Figure 3.17: Seasonal Differencing Distribution, Masinga

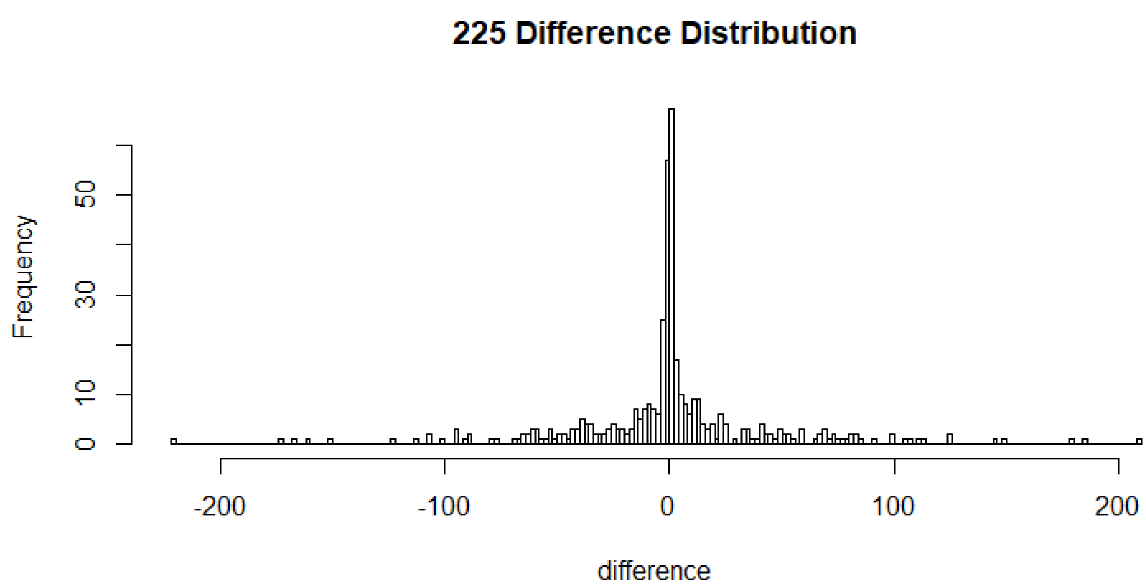


Figure 3.18: Cumulative Distribution of Seasonal Differencing, Matungulu

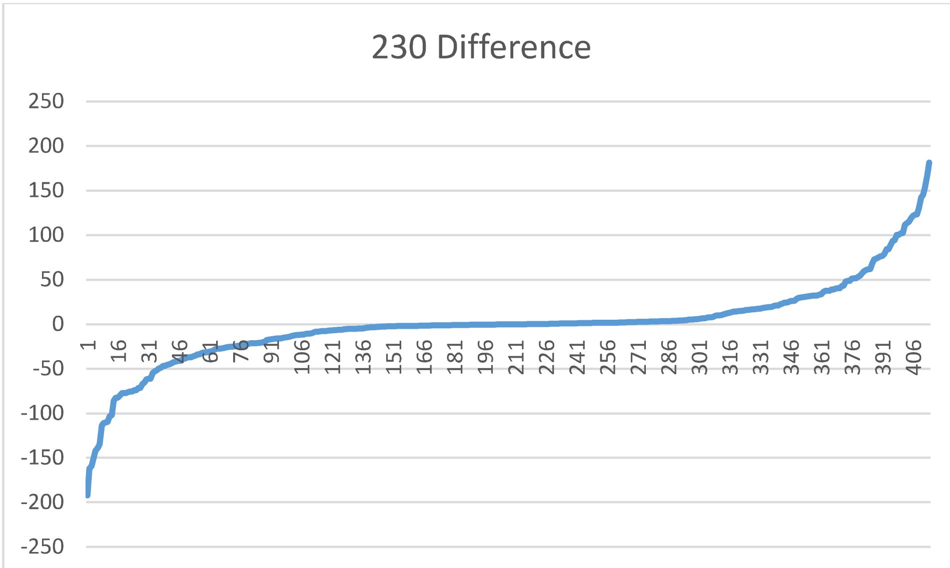


Figure 3.19: Seasonal Differencing Distribution, Matungulu

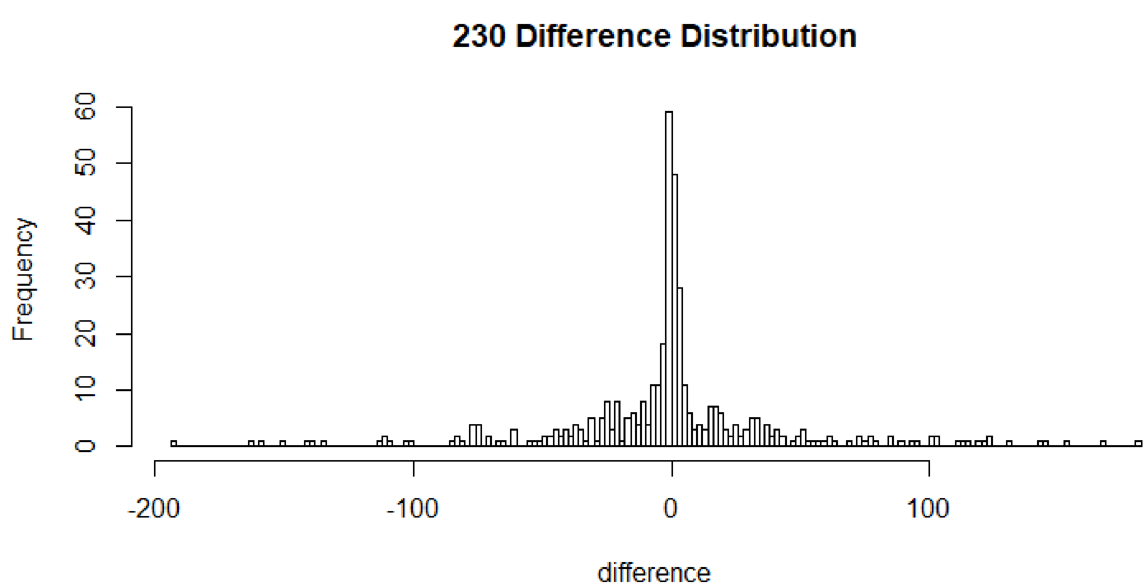


Figure 3.20: Cumulative Distribution of Seasonal Differencing, Kalama

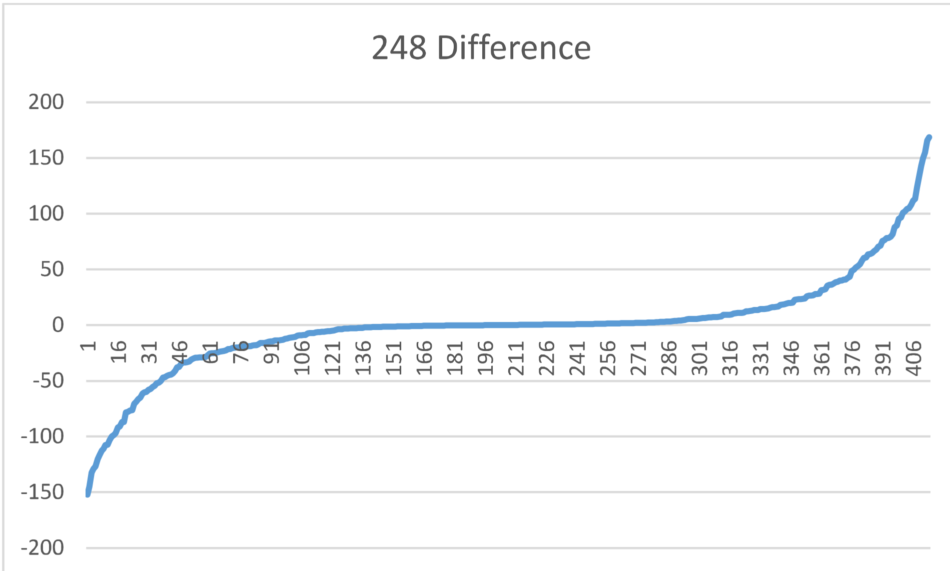


Figure 3.21: Seasonal Differencing Distribution, Kalama

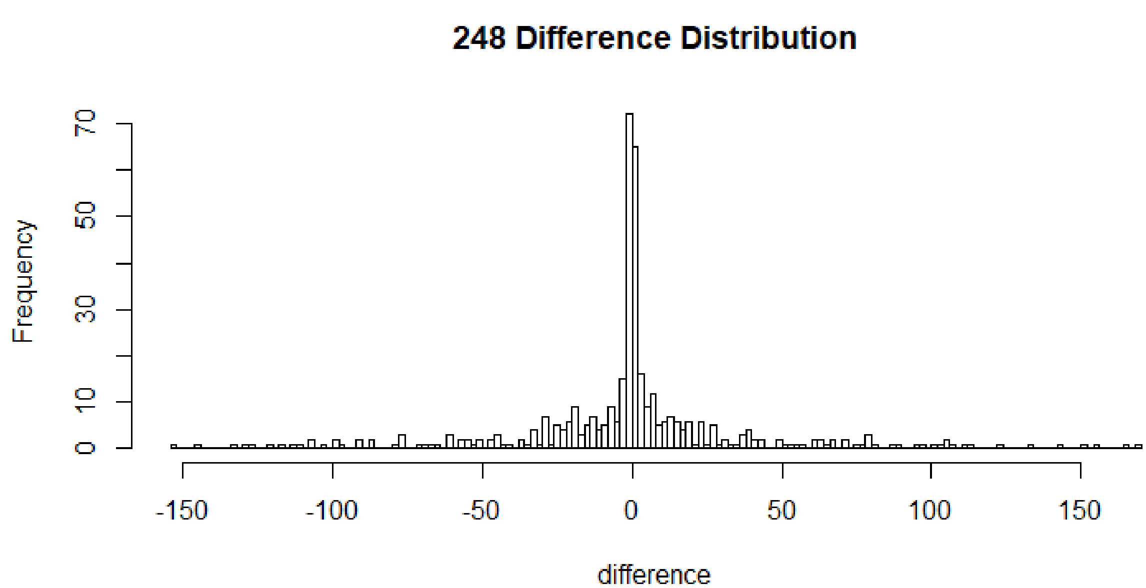


Figure 3.22: Cumulative Distribution of Seasonal Differencing, Kathiani

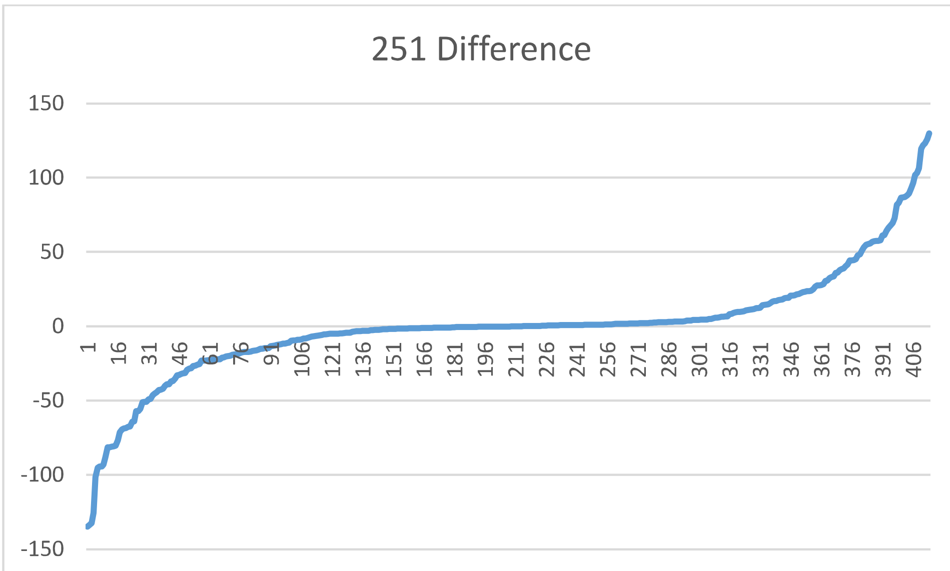


Figure 3.23: Seasonal Differencing Distribution, Kathiani

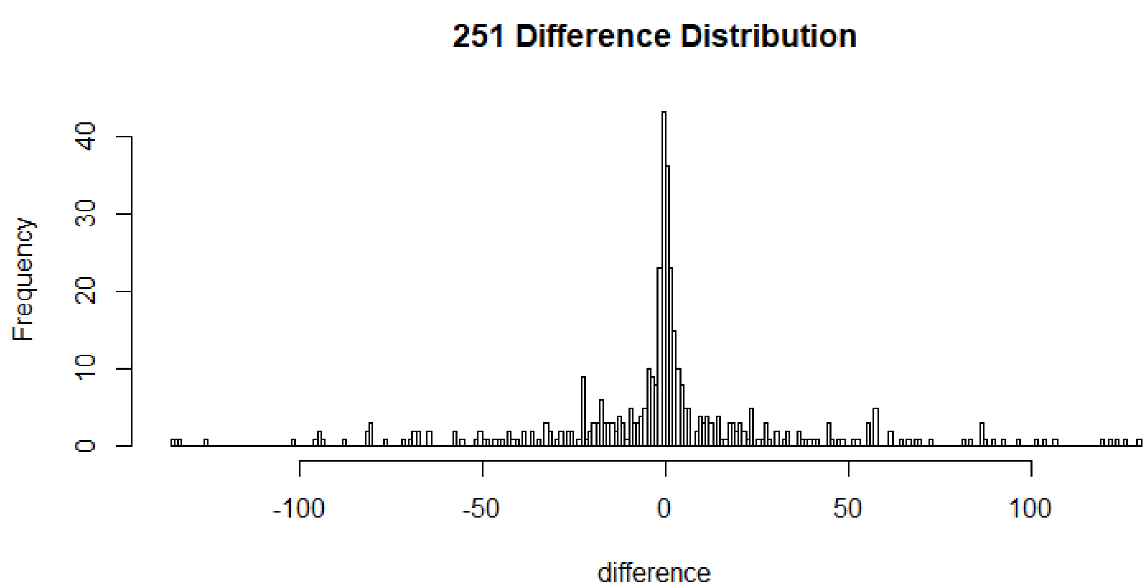


Figure 3.24: Cumulative Distribution of Seasonal Differencing, Mwala

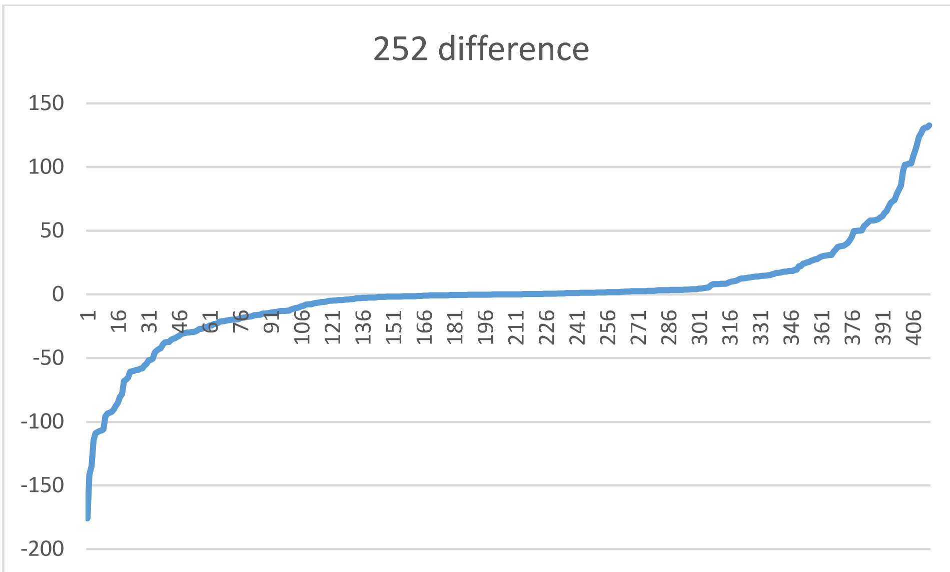


Figure 3.25: Seasonal Differencing Distribution, Mwala

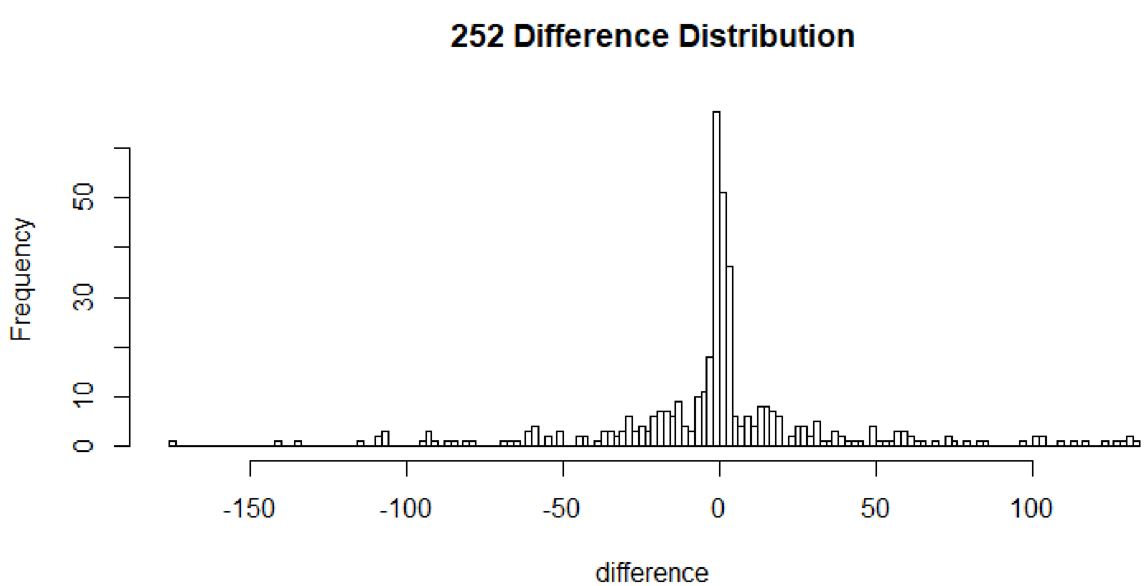


Figure 3.26: Cumulative Distribution of Seasonal Differencing, Kangundo

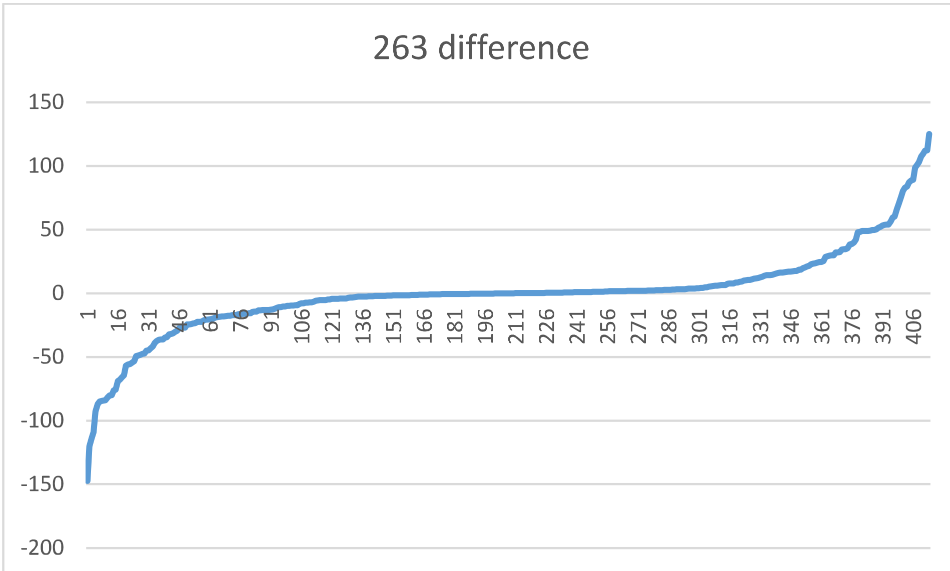


Figure 3.27: Seasonal Differencing Distribution, Kangundo

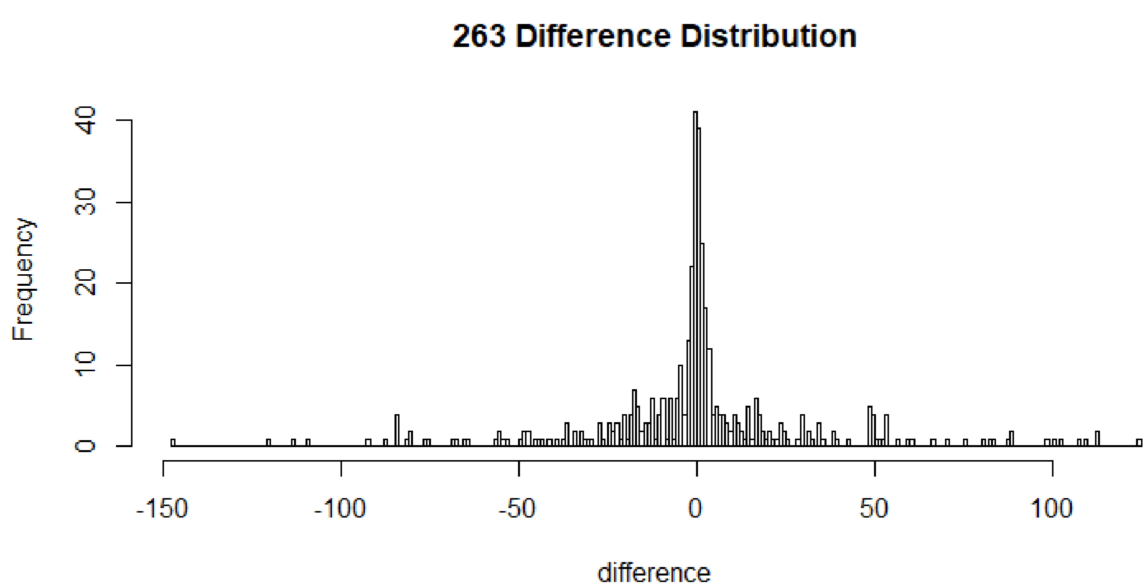


Figure 3.28: Cumulative Distribution of Seasonal Differencing, Ndithini

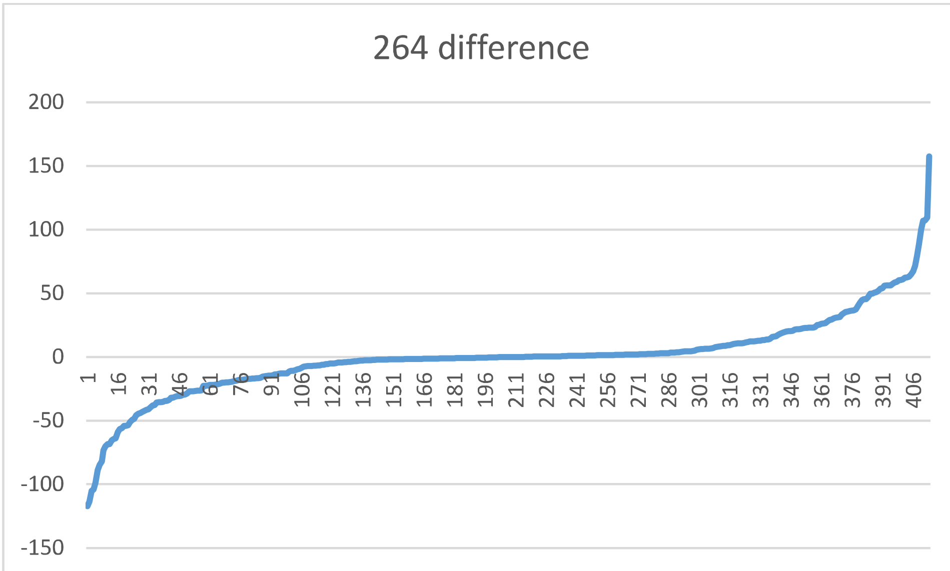


Figure 3.29: Seasonal Differencing Distribution, Ndithini

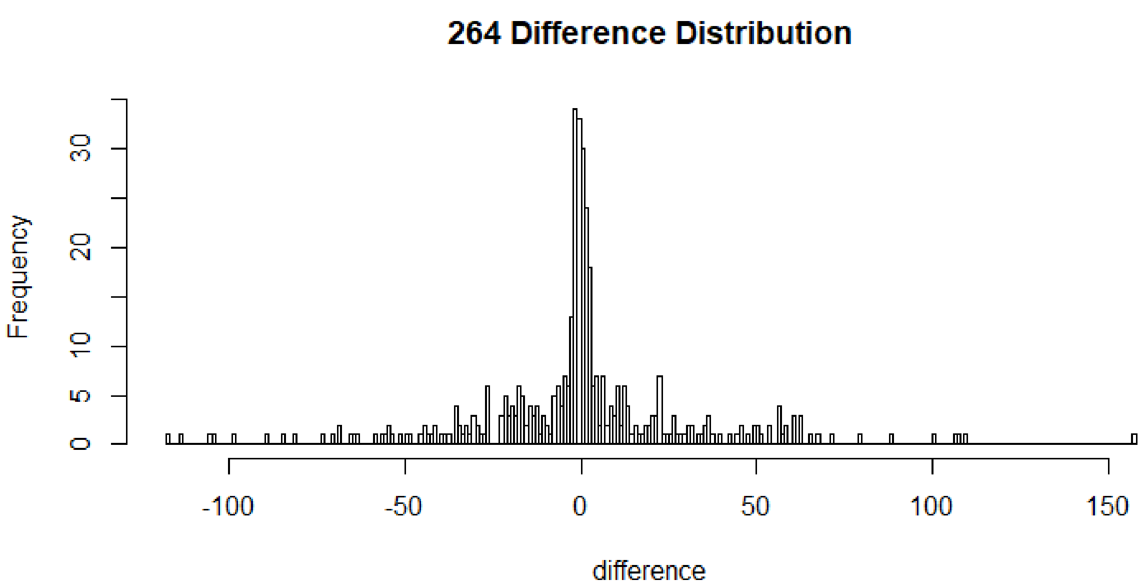


Figure 3.30: Cumulative Distribution of Seasonal Differencing, Mavoko

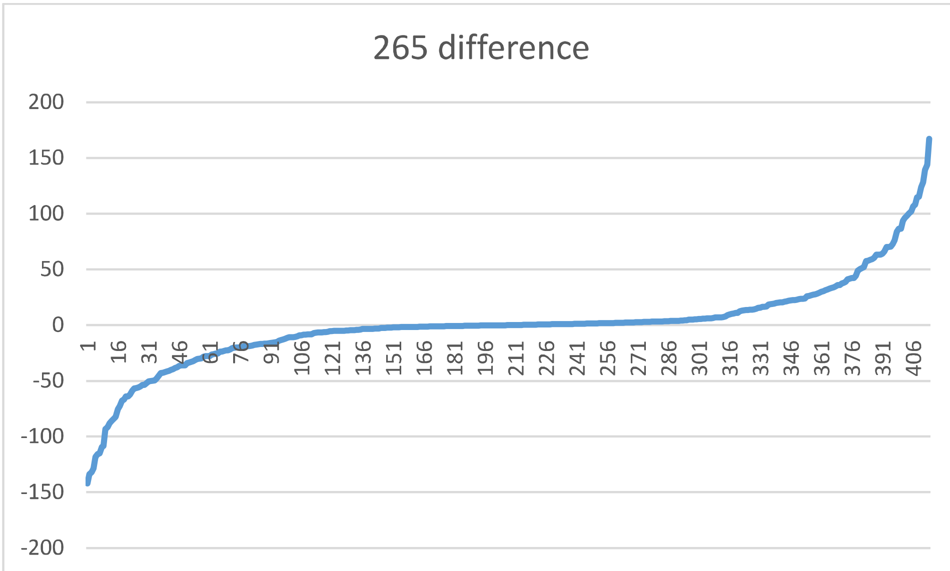
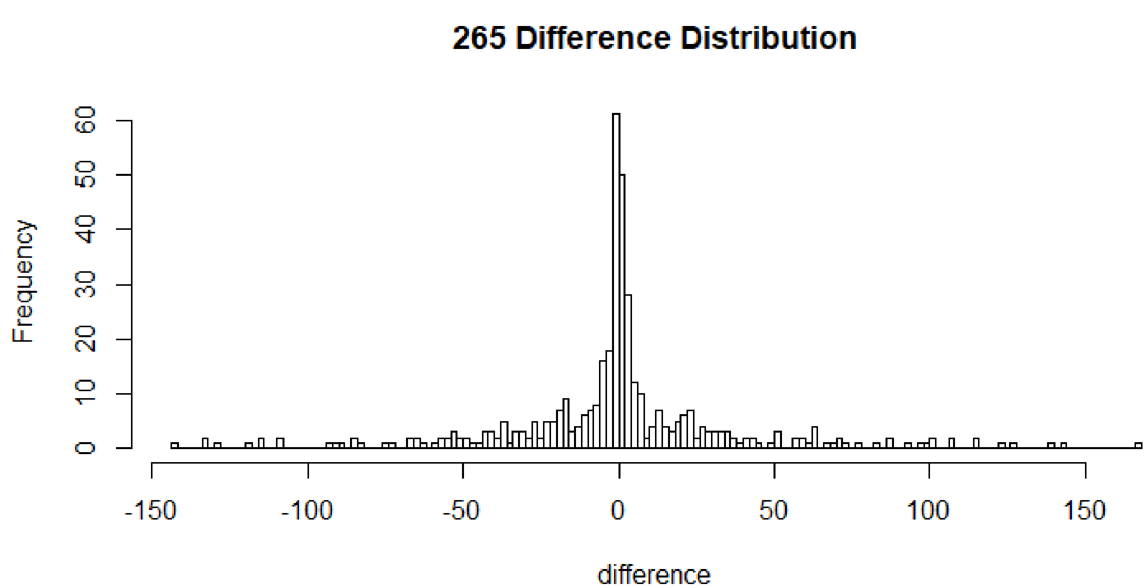
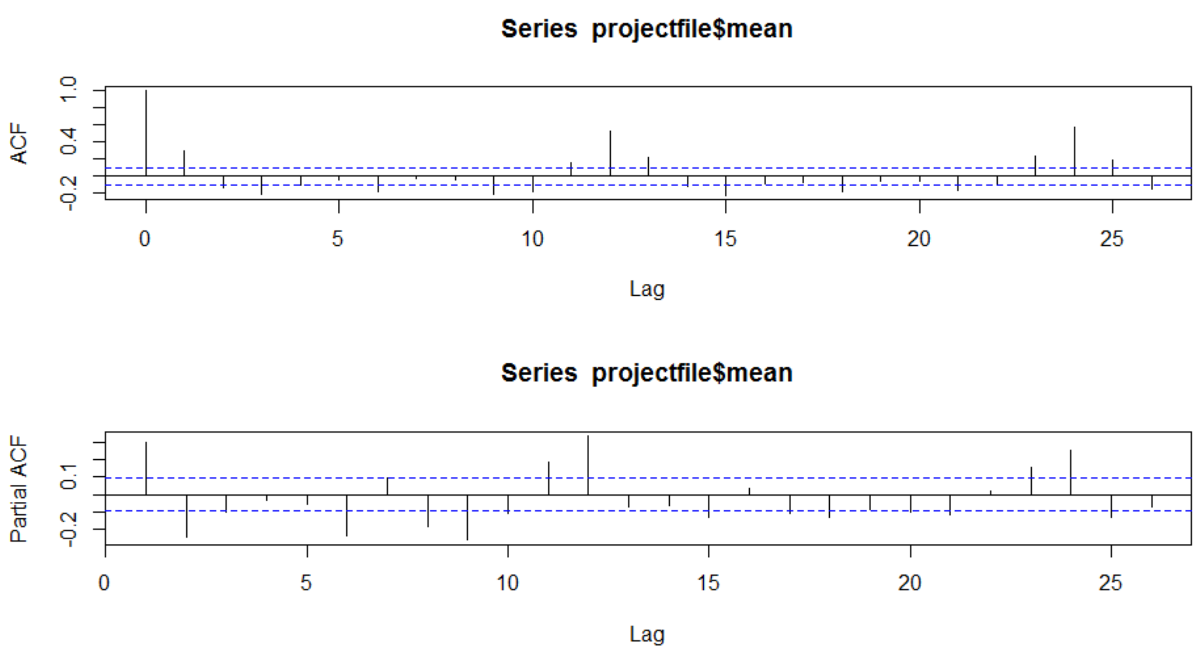


Figure 3.31: Seasonal Differencing Distribution, Mavoko



Meanwhile, the ACF and PACF of original data in Figure 3.32 shows that there is an obvious seasonal pattern within the rainfall. Spikes show up in 1, 12, 24 in both ACF and PACF graph, which confirms that rainfall pattern is seasonal with a frequency of 12. Within in a period, period 1, 2, 3 and 4 are significant, which composes the non-seasonal part. In the following parts, I am going to fit my model with this pre-processed data set. For the moving average (MA) part, we select parameters by trying from 0, and 1.

Figure 3.32: ACF and PACF Plot of Original Data



Chapter 4: Results

The goal of this chapter is to decide the coefficient and parameters for SARIMA model in 11 plots in Kenya. AIC will be used here to select the best-fitted model in each area. To make sure the quality of the model, corresponding check on residuals will be applied to see whether current models capture all patterns within the rainfall dataset.

4.1 SARIMA model and parameters selection

4.1.1 Model selection criteria

The goal of using the time series data is to predict the values accurately in the future. There are some standards that can be used to determine the best fitting model that is not over-fitted. Namely, we can use in-sample criteria and out-of-sample criteria.

The In-sample criteria is calculated using the forecast errors. Mean Squared Error is equal to:

$$MSE = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2, \quad (1)$$

$$\hat{u}_s = y_s - \hat{y}_s^{(s-1)} \quad (2)$$

These are called one-step ahead forecast errors. y_s is the observation from the real series, and $\hat{y}_s^{(s-1)}$ comes from prediction of the previous value. This is then used to compute the Akaike information criterion (AIC):

$$AIC = \ln(MSE) + 2\frac{p}{T} \quad (3)$$

In the function above, p is the number of parameters in the model. AIC penalizes models that add parameters that do not reduce the MSE very much. Generally speaking, a small AIC indicates a good model.

Out-of-sample criteria estimate the predictive accuracy of the model by splitting the series in two parts. One is training set, which is use to generate the time series function, and another is test set, which is used to validate the accuracy of the prediction function generated from training set.

Forecast errors can be calculated as:

$$e^{h}_{S+h} = y_{S+h} - \hat{y}^{(S)}_{S+h}. \quad (4)$$

Root Mean Squared Error is:

$$RMSE_h = \sqrt{\frac{1}{T-S-h+1} \sum_{t=S+h}^{T} (e^h_t)^2}. \quad (5)$$

T is the total size of the series. S is the size of the training set. h is the size of test set. Out-of-sample criteria approach may require the model to be re-estimated many times by splitting a different training set and test size. The size of two sets is critical, and all should be large enough to support the validation process.

Using the model created in the previous part, a set of predicted values is generated.

$$\hat{y}^{(T)}_s = E[Y_s | y_1, y_2, \ldots, y_T] \quad (6)$$

It is then possible to construct a prediction interval for some degree of confidence, such as 95%:

$$P(\hat{y}^{(t)}_{t+h} - 2\sigma_h \leq y_{t+h} \leq \hat{y}^{(t)}_{t+h} + 2\sigma_h) = 95\%. \quad (7)$$

Therefore, we can conclude that $\hat{y}_{t+h}^{(t)} \pm 2\sigma_h$ is the 95% prediction interval.

4.1.2 Application

I used AIC to help select the best-fitted model. A seasonal ARIMA model with a frequency of 12 was chosen for the basic form of the model. In addition, looking at the ACF graph above, notice that the spikes are only significant until 4 lags. The 5th spike is within the dashed line. It implies an AR model with order of 4 is appropriate.

By apply auto.arima function in R, from the 'forecast' package, the best-fitted model is generated automatically, trying different specifications up to order 5 on each parameter. The results show a seasonal ARIMA(4,0,1)(1,0,0)[36] model with AIC=3318.83. Each month has 3 observable data points; therefore, the frequency here is 36. In addition, d equals to 0 here and it confirms that there is no trend within the series, and differencing is not necessary.

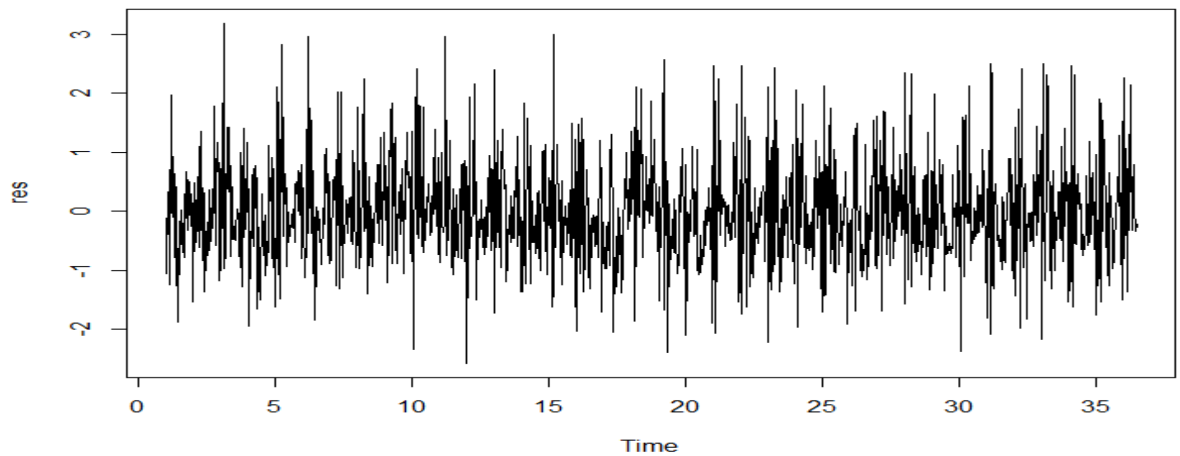
Table 4.1: SARIMA Model Parameters and Coefficient

ARIMA(4,0,1)(1,0,0)[36]	AR1	AR2	AR3	AR4	MA1	SAR1
Coefficient	1.4829	-0.3754	-0.0430	-0.1112	-0.9800	0.3159
s.e.	0.0303	0.0511	0.0502	0.0284	0.0086	0.0334

4.1.3 Residual check

After deciding the parameters of the model, the next step is to apply residual diagnostics. A good model should have normally distributed. The figure below shows the distribution of residuals.

Figure 4.1: Plot of Residuals



After performing Ljung-Box test, with a lag equals to 36 and difference equals to 0, the calculated p-value is 1.423e-10. The test result indicate that the residuals follow a normal distribution.

Ljung-Box test can be written as:

H_0: The data are independently distributed

H_a: The data are not independently distributed; there is serial correlation

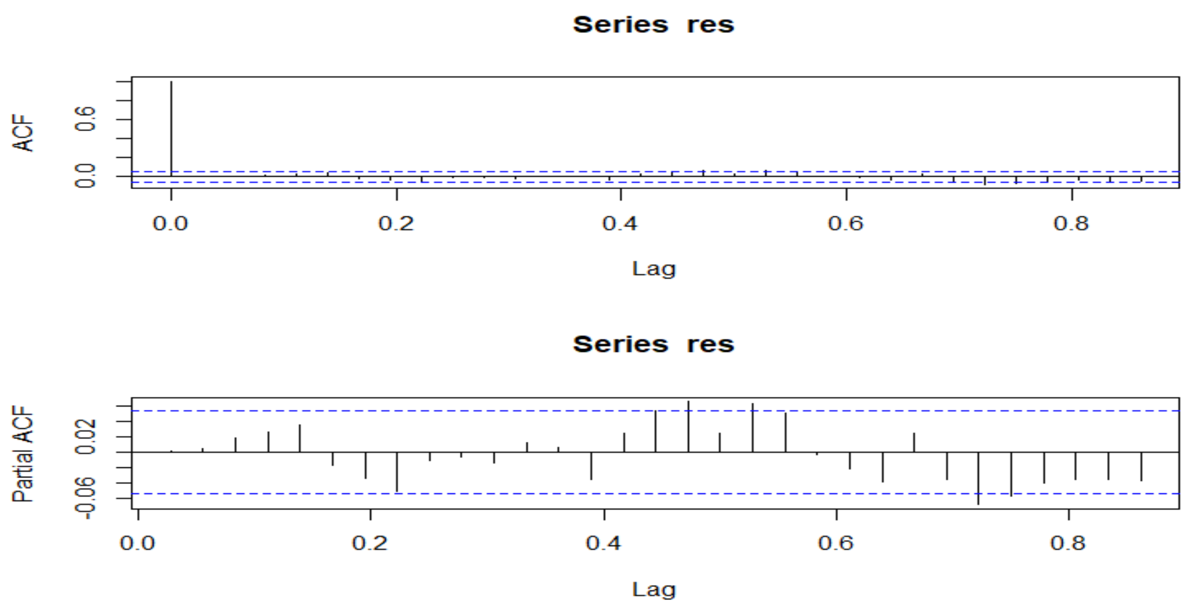
$$Q = n\,(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

$\hat{\rho}_k$ is the sample correlation at lag k, and h is the number of lags being tested.

$$Q > \chi^2_{1-\alpha,h}$$

In addition, after applying ACF and PACF to residuals, I find that residuals are uncorrelated, because for ACF figure, it only has spike at 0, and after the first, there is no obvious correlation among residuals. Since all of the necessary conditions are satisfied, this seasonal ARIMA model is a good fit for further forecasting.

Figure 4.2: ACF & PACF of Residuals



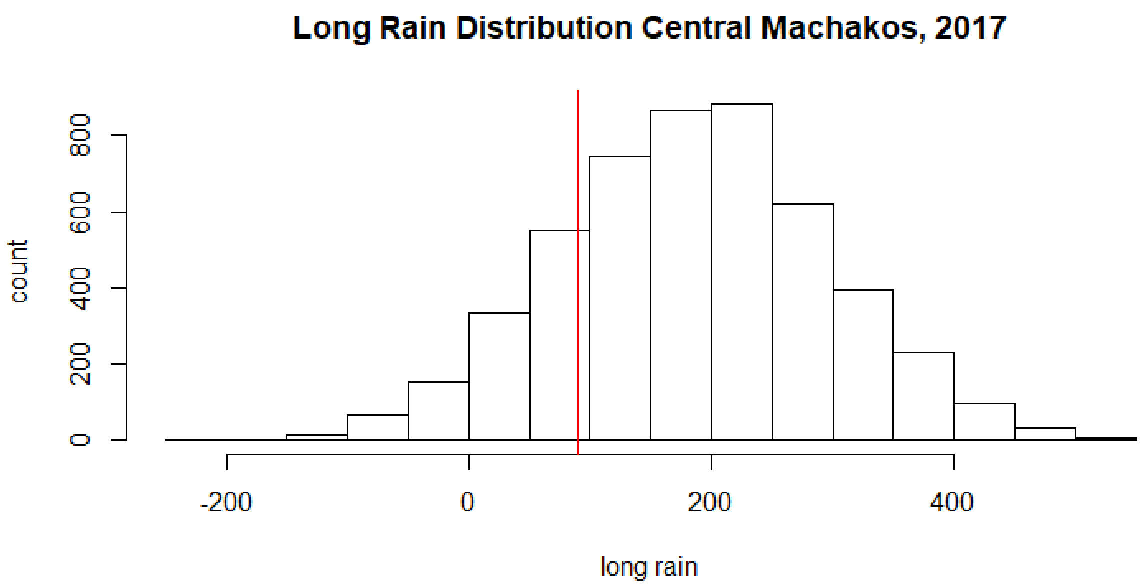
If the distribution of residual is not white noise, for example, if it has a serial pattern within the residuals, then a linear time series model doesn't fit this situation well. Non-linear time series models, such as GARCH (Generalized AutoRegression Conditional Heteroscedasticity)and ARCH (Autoregressive Conditionally Heteroscedasticity will be considered have to be considered to capture these nonlinearities.

4.1.4 Forecast and lower 20% quantile of simulation

After selecting the best-fitting model for the rainfall, I use SARIMA model with parameters and coefficients to generate predicted rainfall values for the future. I run 5000 simulations, which provides 5000 possible paths of the rainfall in the following observation Figure. The original dataset ends on 26th June 2016, and therefore the goal is to predict the long rain season from 16th Oct. 2016 to 15th January 2017.

16th Oct. 2016 is the 11th prediction after 26th June 2016, and 15th January 2017 is the 20th prediction of the forecast. The sum of these 10 observations is the rainfall total for the long rain season. There should be 5,000 totals long rain predictions according to the times of simulation (n=5000), and we can create a distribution according to these 5,000 estimates. Figures below indicate the distribution for Central Machakos. Figure of all areas can be found in later parts.

Figure 4.3: Distribution of Simulation and 20% lower quantile, Central Machakos



The histogram above shows the distribution of rainfall totals in the long rain season, and the vertical line shows the lower 20% quantile of the rainfall, which is 90.7 mm. Therefore, this number is the trigger that should be used for the weather-index insurance. However, in the distribution of the simulation, there are some negative predicted rainfalls, which is inappropriate for real life rainfall situation.

In reality, the climate system is very complicated, and many mutual interactions should be considered. In future work, a relationship between the rainfall and temperature could be explored to construct a more accurate model. The table below shows the results generated by both the pert distribution method and the SARIMA method, which is the number of lower 20% quantile of the simulated distribution.

4.2 Results Comparison

Table 4.2: Cumulative Rainfall Trigger table (mm)

Place	Central Machakos	Yathui	Yatta	Masinga	Matungulu	Kalama	Kathiani	Mwala	Kangundo	Ndithini	Mavoko
Pert 20%	120.5	141.9	145.2	186.7	143.2	137.2	118.9	132.2	109.5	105.9	122.1
TS 20%	90.7	171.8	235.3	280.2	134.5	93.4	115.7	164	134.6	103	59
Ratio (Pert/TS)	75.27%	121.07%	162.05%	150.08%	93.92%	68.08%	97.31%	124.05%	122.92%	97.26%	48.32%

Average ratio for the table above is 1.08, and standard deviation is 0.33. A simple t-test around 1 can be written as:

$$t=\frac{1.08-1}{0.33}=0.24$$

According to the result above, we fail reject the hypothesis that ratio equals to 1 at 0.1 significance level.

The triggers provided by two methods are generally different. Results provided by SARIMA can be either bigger or smaller than results provided by the pert distribution.

Figure 4.4: All Triggers Comparison Between Two Methods

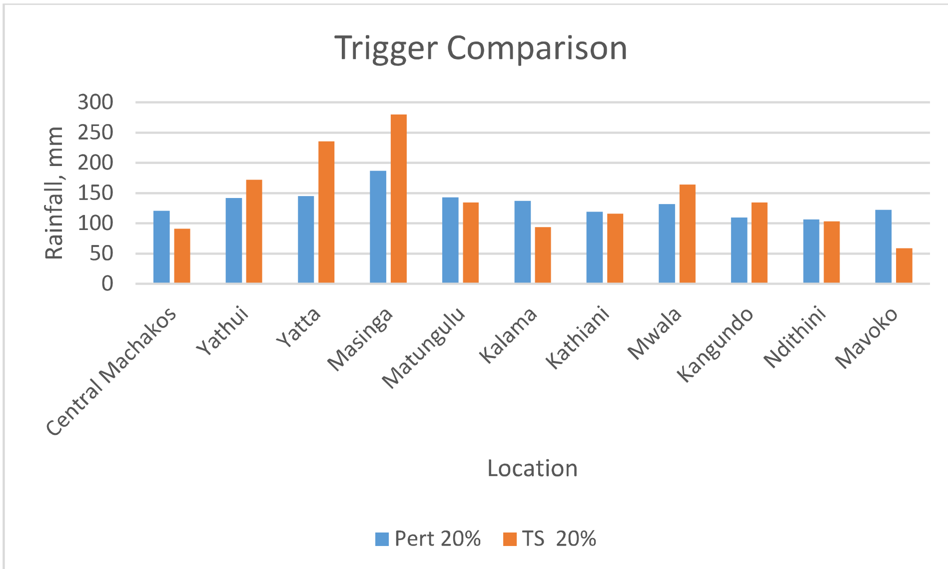


Figure 4.4 also compares the trigger for two methods. There is not a very stable correlation between two methods. Results provided by SARIMA method is bigger than Pert Distribution in Yathui, Yatta, Masinga. Results provided by SARIMA method is smaller than Pert Distribution in Central Machakos, Kalama, and Mavoko. The rest parts have pretty close triggers.

According the trigger comparison graph above, it becomes clear that SARIMA method fails to place a lower bound on rainfall, because forecast results generated by SARIMA method have negative value, which distort the actual distribution of the rainfall. Therefore, projected rainfall distribution generated by SARIMA method under this “negative rainfall exists” condition is more symmetric and less skewness than actual rainfall distribution, which has a non-negative bundle.

The Graphs below show results generated by the pert distribution and the SARIMA model. Nonetheless, the marginal deviations of SARIMA method from the historical pert distributions are too great to be considered as the trigger of the insurance.

Figure 4.5: Trigger Comparison Between Two Methods, Central Machakos

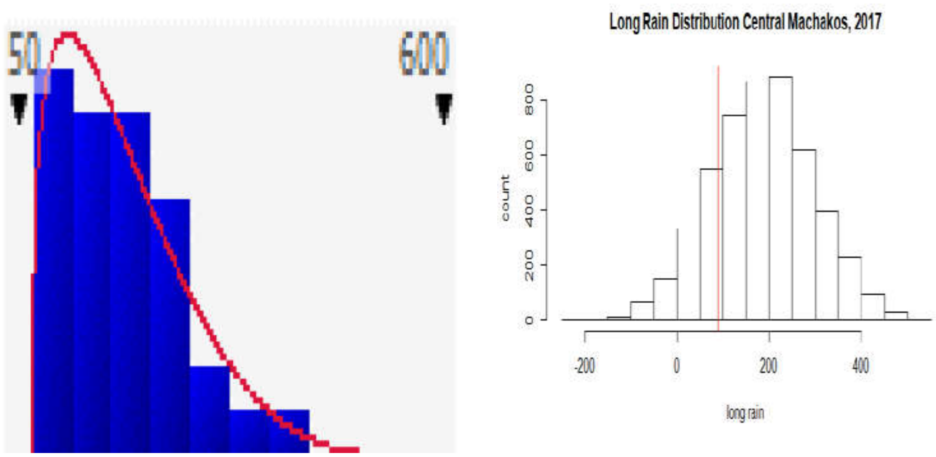


Figure 4.6: Trigger Comparison Between Two Methods, Kalama

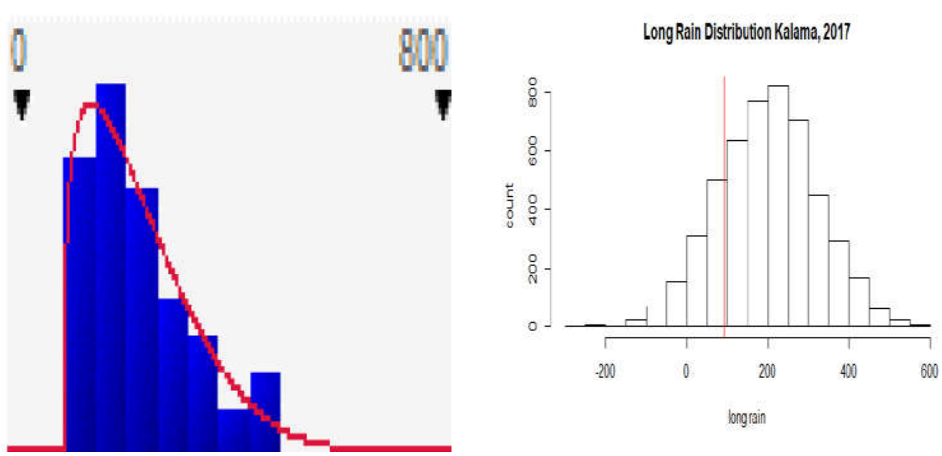


Figure 4.7: Trigger Comparison Between Two Methods, Kangundo

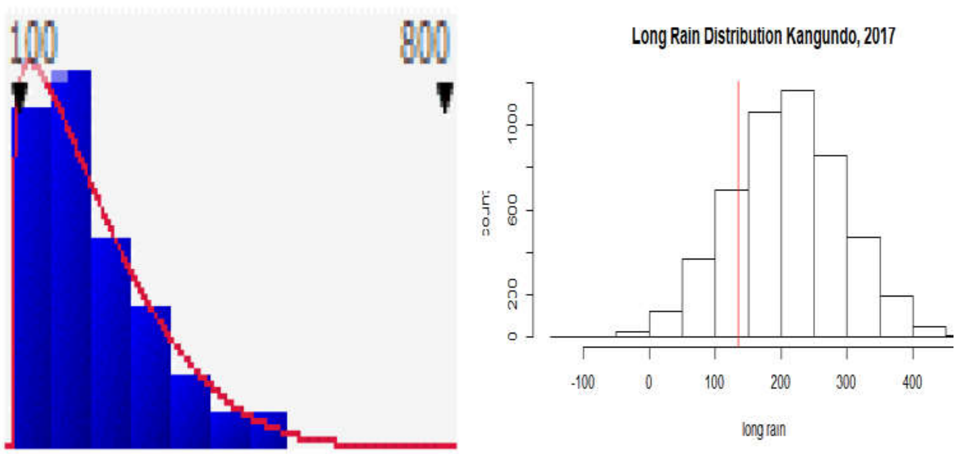


Figure 4.8: Trigger Comparison Between Two Methods, Kathiani

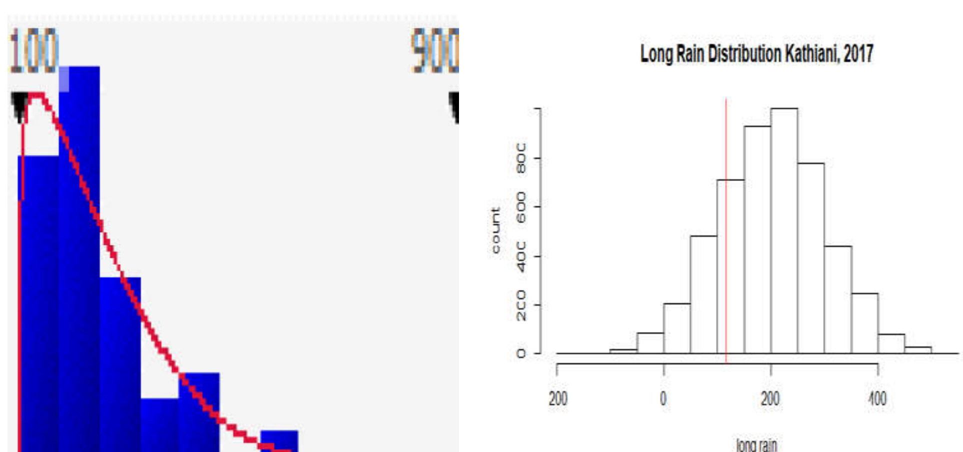


Figure 4.9: Trigger Comparison Between Two Methods, Masinga

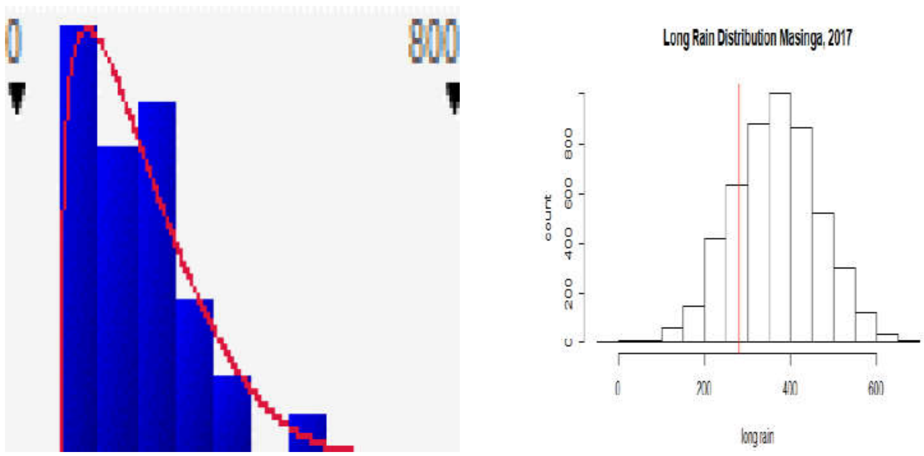


Figure 4.10: Trigger Comparison Between Two Methods, Matungulu

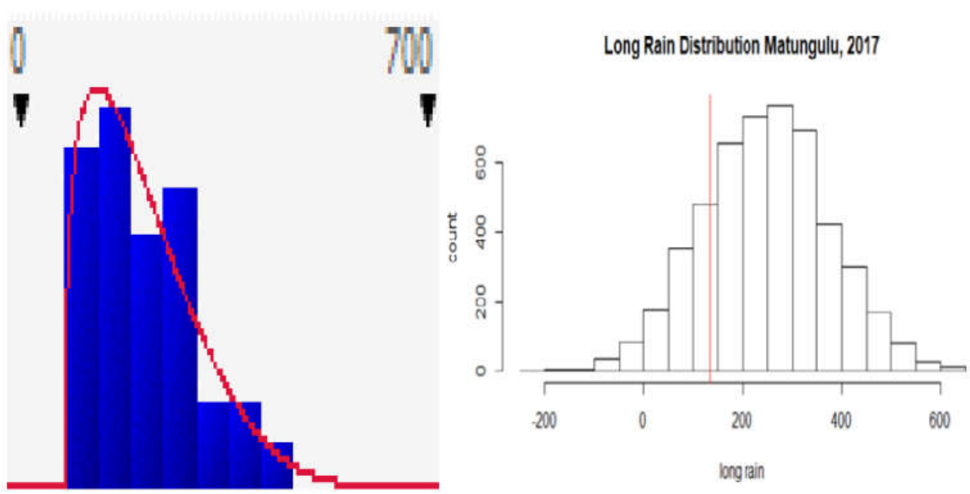


Figure 4.11: Trigger Comparison Between Two Methods, Mavoko

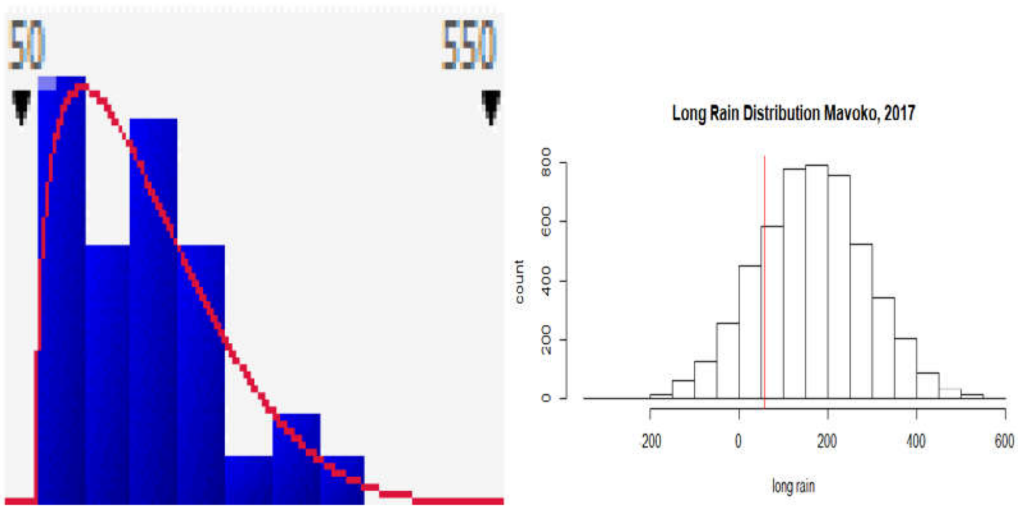


Figure 4.12: Trigger Comparison Between Two Methods, Mwala

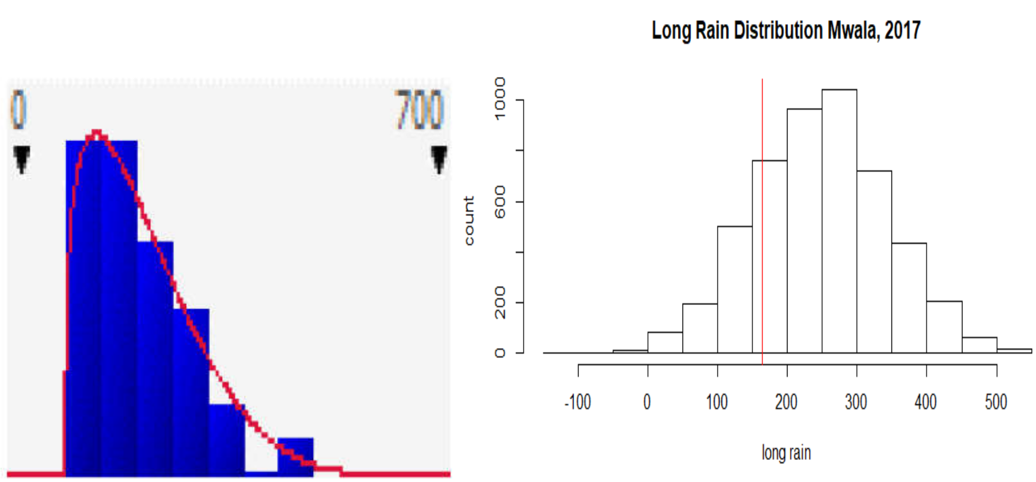


Figure 4.13: Trigger Comparison Between Two Methods, Ndithini

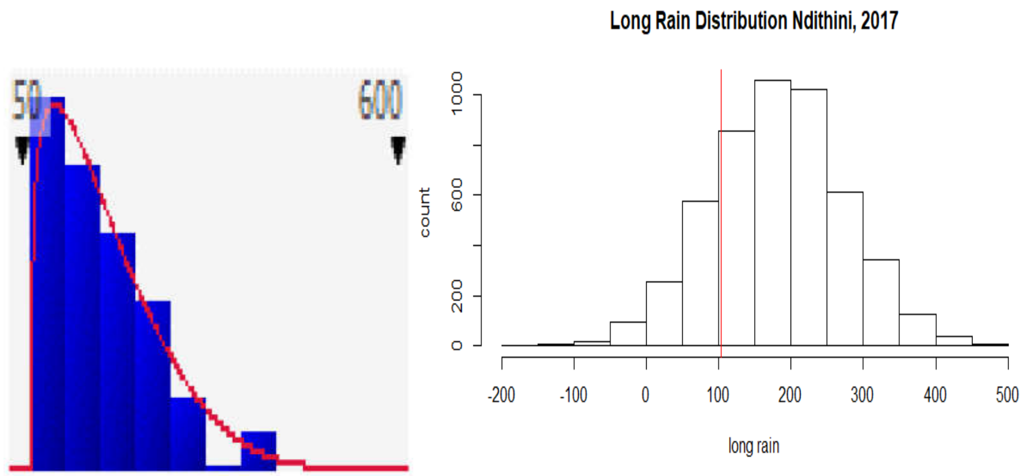


Figure 4.14: Trigger Comparison Between Two Methods, Yathui

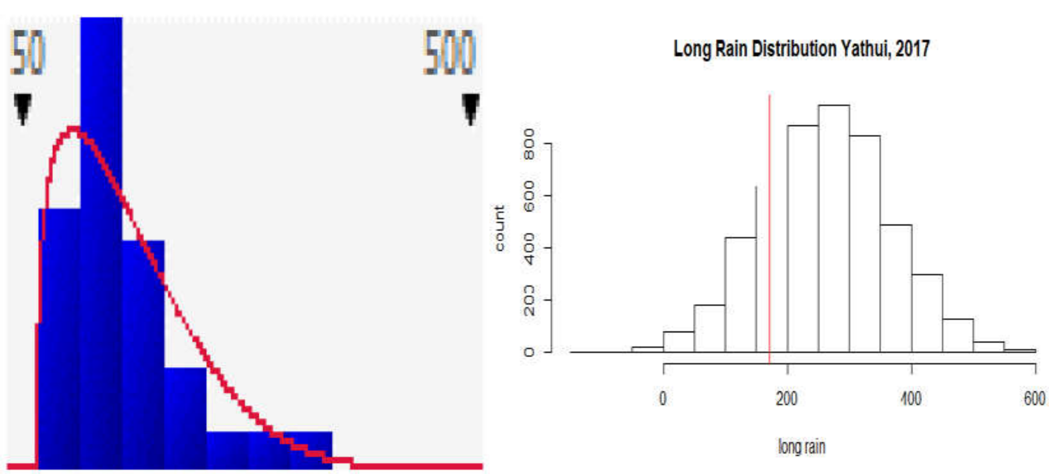
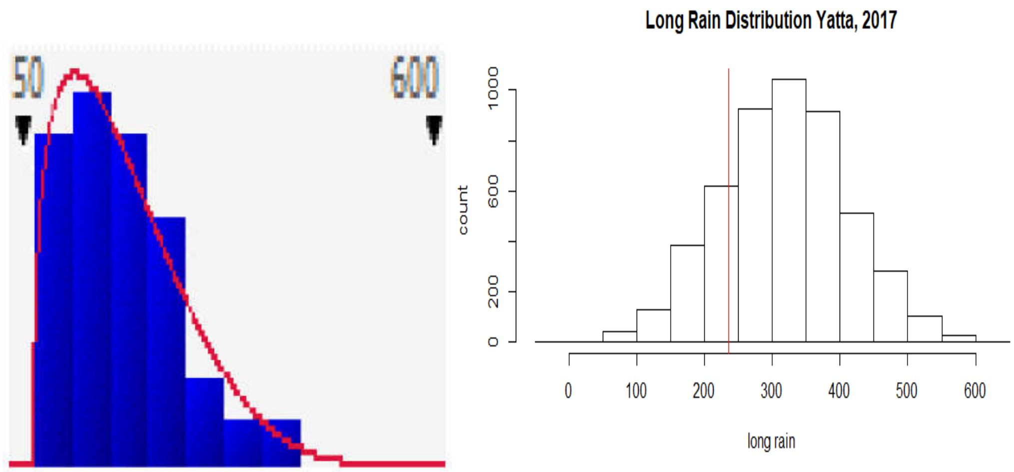


Figure 4.15: Trigger Comparison Between Two Methods, Yatta



Chapter 5: Conclusion

Risk-contingent credit is a flexible option for small household in Kenya to solve weather risks, such as drought efficiently. Weather index insurance is an application of RCC. The performance of underlying assets, which is rainfall determines the payoff of the insurance. After considering real rainfall, RCC connects the performance of underlying assets and payoff function of the insurance dynamically, which guarantees the sustainable cash flow for farmers. Therefore, even farmers suffer from drought, they can still get paid and continue to begin the production for the following years.

The purpose of this thesis was to capture patterns embedded historical rainfall, and therefore make prediction on future rainfall. To achieve this purpose, I use two methods to simulate rainfall pattern, which are Pert Distribution and SARIMA model. After comparing results generated two methods, I conclude that Pert Distribution method is more appropriate to simulate rainfall and provide trigger than SARIMA model. In the prediction of rainfall, the SARIMA model will generate negative values, which is not consistent with the reality. Those negative numbers may distort the accuracy of the prediction. The advantages of the pert distribution are that all numbers in the distribution are positive, and that the trigger won't change from year to year.

The methodology of time series analysis was used to forecast rainfall and find the corresponding confidence interval. The lower bound of the confidence interval can be used as a trigger for the weather-index insurance premium calculation. Both the insurance companies and farmers can use this as a reference when planning. While the pert distribution method works too, the flexibility and responsiveness of SARIMA to new data gives insurers a new tool to use in the determination of the trigger for weather-index insurance

For the SARIMA model, insurance companies will update benchmarks over time. Clients of the insurance may be skeptical about frequently updated triggers. They may think that insurance companies are cheating during the process because they are manipulating triggers to be favorable to the insurance companies themselves. For example, for this weather index-insurance, farmers may think that insurance companies are intentionally decreasing the value of the trigger, so issuers can reduce the probability of having to pay compensation. A stable trigger can therefore be valuable.

The launch of this risk contingent credit embedded financial instruments provides more options for farmers to hedge risks in production, and it is a new opportunity for industry to earn profit in a new market. For those insurance companies, they can construct a portfolio which includes farmers from different areas to diversify risks.

The base of potential customers for the insurance is huge, and the insurance will benefit all farmers who are exposed to weather risks. Unpredictable weather events will have a high chance to put previously self-sufficient farmers below the threshold of the poverty trap. Once farmers fall into poverty traps, it is extremely difficult for them to get out of the trap without help from the outside. The role of weather-index insurance is to protect these farmers from weather shocks that may push them into a poverty trap. Weather-index insurance is application on credit bundled with insurance, which makes agricultural insurances more accessible to small households in Kenya. When insured farmers are inevitably affected by weather shocks and their production level goes below the poverty trap threshold, they can still get capital and keep their farm sustainable for following crop seasons.

Farmers who are slightly above the threshold will benefit the most from this risk contingent credit embedded product. They are fragile to catastrophic shock and index insurance can help them to solidify their level of production. Index-insurance is not panacea. For farmers who are already in low-production equilibrium, this insurance can't prevent them from falling into poverty trap.

In the future, more parameters which can capture local weather patterns will need to be included into this risk contingent credit pricing model, and repayment function will also be adjusted accordingly to satisfy different cases. For example, besides rainfall, temperature is another important factor for the growth of crops. If the premium function can incorporate more independent variables, the accuracy of the model will be improved.

With accurate rainfall pattern forecasts, insurance companies can design fairly-priced risk contingent credit embedded products that farmers find worth buying. Ideally, a practical insurance product should capture the pattern of insured goods accurately. It should also be easily observable and objective.

Risk contingent credit embedded financial instrument is a promising instrument for farmers, and it has natural advantages over traditional claimed insurance. The objectivity of index-insurance avoids moral hazard completely. Since it provides a universal benchmark, which also eliminate the need for the insurance company to evaluate loss and indemnity case by case. Every characteristic of index-insurance mentioned above will make this financial instrument more affordable and accessible to farmers, especially those in rural areas.

How to find right indicators to represent the actual suffering of farmers becomes a very important question. There are cases in which insurance cannot compensate the loss borne by farmers accurately. There might be other important factors that also influence production; however, the mechanism of the credit and insurance doesn't include those equally important factors, which will damage the efficiency of the insurance.

Besides influential indicators, correlations among indicators are very important and difficult to deal with. Even if researchers learn about all of the meaningful factors, problematic correlations among indicators will also hurt the value of insurance. For example, correlation between high temperature and possibility of hitting drought should be different from the correlation between moderate temperature and possibility of hitting drought. Therefore, with the introduction of more weather indicators, such as temperature, the correlation among those indicators can't be captured sufficiently by a constant, and correlation should be changeable according to different combination of weather indicators.

In future, more research should focus on the possible correlation simulation, either within the same rainfall season or between the long rain and the short rain seasons. The copula method will be a powerful tool to solve this issue. Finding all of the relevant indicators and their correlations to one another is an essential role for future research.

In addition, correlations among different parameters, such as precipitation, and temperature will be taken into consideration. Copulas can be a way to simulate changeable correlation between parameters. Copulas are used to describe the dependence between random variables. For example, the correlation between temperature and rainfall is not constant- correlation between low rainfall and high temperature is much stronger than correlation between medium rainfall and medium temperature. Traditional constant correlation between random variables doesn’t capture the complexity of climate relationships.

References:

Barnett, B.J., Christopher Barrett, and J. Skees (2008), "Poverty Traps and Index-based Risk Transfer Products," World Development 36: 1766-1785.

Barrett C.B., Barnett B.J., Carter M.R., Chantarat S., Hansen J.W., Mude A.G., Osgood D., Skees J.R., Turvey C.G. and Ward M.N. (2007) Barrett C.B., Barnett B.J., Carter M.R., Chantarat S., Hansen J.W., Mude A.G., Osgood D., Skees J.R., Turvey C.G. and Ward M.N. (2007)

Binswanger-Mkhize, H., "Is there too much hype about index-based agricultural insurance?," mimeo 2011.

Boucher, S.R., Carter, M.R. and Guirkinger, C. (2008), "Risk rationing and wealth effects in credit markets: theory and implications for agricultural development", American Journal of Agricultural Economics, Vol. 90 No. 2, pp. 409-423.

Carter, M.R., and C.B. Barrett. 2006. The Economics of Poverty Traps and Persistent Poverty: An Asset-Based Approach. Journal of Development Studies 42(2):178–199

Carter et al., 2007 M. Carter, F. Galarza, S. Boucher. Underwriting area-based yield insurance to crowd-in credit supply and demand, Giannini Foundation of Agricultural Economics (2007) Working Paper, UC Davis 07-003

Chandra S. Kumar, Calum Turvey, Jaclyn D. Kropp. 2013. The impact of credit constraints on farm households: Survey results from India and China, Applied Economic Perspectives and Policy, Volume 35, 508-527.

Jensen et al., 2017 Nathaniel Jensen, Christopher Barrett, Andrew Mude Cash transfers and index insurance: a comparative impact analysis from northern Kenya J. Dev. Econ., 129 (2017), pp. 14-28

Jensen, N., Mude, A., & Barrett, C.B. (2014). How Basis Risk and Spatiotemporal Adverse Selection Influence Demand for Index Insurance: Evidence from Northern Kenya.

Leslie J. Verteramo-Chiu, Sivalai V. Khantachavana, Calum G. Turvey, (2014) "Risk rationing and the demand for agricultural credit: a comparative investigation of Mexico and China", Agricultural Finance Review, Vol. 74 Issue: 2, pp.248-270

Makaudze, E. (2012), Weather Index Insurance for Smallholder Farmers in Africa: Lessons Learnt and Goals for the Future African Sun Media.

Michael T. Norton, Calum Turvey, Daniel Osgood. 2012. Quantifying spatial basis risk for weather index insurance. The Journal of Risk Finance 14:1, 20-34.

Norton, M. T., Turvey, C., & Osgood, D. (2012). Quantifying spatial basis risk for weather index insurance. Journal of Risk Finance, 14(1), 20-34.

Poverty Traps and Climate Risk: Limitations and Opportunities of Index-Based Risk Financing. IRI Technical Report No. 07-02. IRI, Columbia University, New York.

Robert, S. David, S. (2010) Time Series Analysis and Its Applications with R Examples, New York, NY: Springe

Shee, A., Turvey, C.G. and Woodard, J. (2015), "A field study for assessing risk-contingent credit for Kenyan pastoralists and dairy farmers", Agricultural Finance Review, Vol. 75 No. 3, pp. 330-348.

Shee, A. and Turvey, C.G. (2012), "Collateral-free lending with risk-contingent credit for agricultural development: indemnifying loans against pulse crop price risk in India", Agricultural Economics, Vol. 43 No. 5, pp. 561-574.

Shee, A, Turvey C.G., You L, "Design and Rating of Risk-Contingent Credit for Balancing Business and Financial Risks for Kenyan Farmers", (2015) design paper

Stiglitz, Joseph E., and Andrew Weiss. "Credit Rationing in Markets with Imperfect Information." The American Economic Review, vol. 71, no. 3, 1981, pp. 393–410. JSTOR

Woodard, J.D. and Garcia, P. (2008a), “Basis risk and weather hedging effectiveness”, Agricultural Finance Review, Vol. 68 No. 1, pp. 99-118.

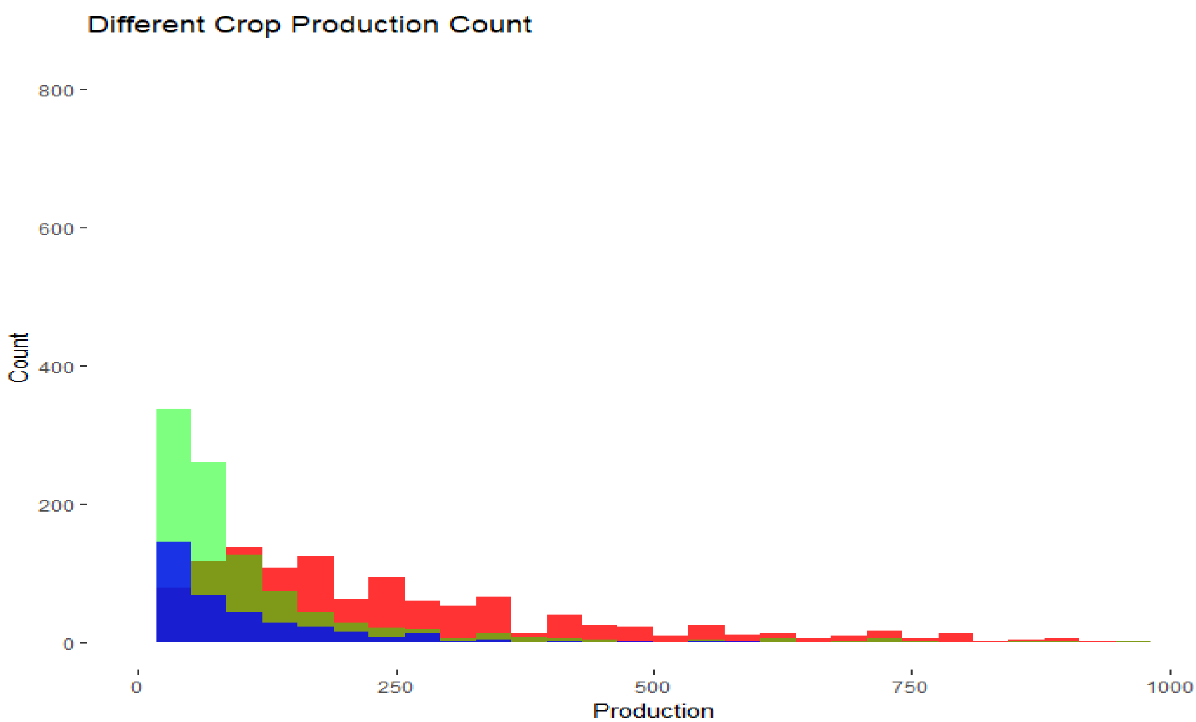
Yucemen, M. S. "Probabilistic assessment of earthquake insurance rates for Turkey." Natural Hazards 35.2 (2005): 291-313.

Appendix:

```{r}
## % dist graph
data=merged_master
uncom=data                                                    %>%
mutate(perc=food_expense/(food_expense+common_expense+uncommon_expense))

a=as.data.frame(uncom$perc)
p1=ggplot(data=a,aes(x=uncom$perc))+geom_histogram()

p1+theme(panel.background = element_blank())+labs(title="Food Expense Percentage
Among Total Budget",x="percentage",y="count")
```

```{r}
##crop production
maize=as.data.frame(data$maize_per_acre)
maizenona=maize[!is.na(maize)]
bean=as.data.frame(data$bean_per_acre1)
beannona=bean[!is.na(bean)]
pea=as.data.frame(data$cowpea_per_acre1)
```

```
peanona=pea[!is.na(pea)]

p2=ggplot(data=crop)+geom_histogram(aes(x=crop$maize_per_acre),fill="red",alpha
=0.8)+
  geom_histogram(aes(x=crop$bean_per_acre1),fill="green",alpha=0.5)+
  geom_histogram(aes(x=crop$cowpea_per_acre1),fill="blue",alpha=0.8)+
  scale_x_continuous(limits=c(0,1000))+theme(panel.background = element_blank())+
  labs(title="Different Crop Production Count",x="Production",y="Count")
```



Chapter 4 code:

```
model <- Arima((data$rain) , order= c(4, 0 ,1), seasonal= list(order = c(1, 0, 0), period = 36), include.mean
```

```
= FALSE)

sims <- 1000

sims_result <- NULL


for(i in 1:sims) {


    foo <- simulate(model, nsim = 20)

    #index 11 is 16-oct-16, index 20 is 16-jan-17

    sims_result <- rbind(sims_result, (foo[11:20]))


 }

 sums <- apply(sims_result, 1, sum)

(sort(sums))[200]


hist(sums,xlab="long rain", ylab="count",main="Long Rain Distribution 2017")

quantile(sums,0.2)

abline(v=quantile(sums,0.2),col="red")
```

Other SARIMA parameters and resulting coefficients:

Plot 213: (Yathui)

```
Series: (n213$`213`)

ARIMA(4,0,0)(1,0,0)[36] with zero mean


Coefficients:

         ar1     ar2     ar3     ar4    sar1

      0.4136  0.1523  0.1139  0.0104  0.2666

s.e.  0.0309  0.0302  0.0301  0.0284  0.0317
```

Plot 214:

Series: (n214$`214`)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	-0.1178	0.4034	0.1021	0.1077	0.6637	0.8590
s.e.	0.7972	0.4532	0.1459	0.1438	0.7918	0.0609

Plot 225: (Yatta)

Series: (n225$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	0.0212	0.4770	0.0085	0.0424	0.5901	0.8856
s.e.	0.0296	0.2224	0.1470	0.1469	0.1037	0.0594

Plot 230: (Masinga)

Series: (n230$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	0.1788	0.3894	0.077	0.0111	0.1744	0.6665
s.e.	NaN	0.0019	NaN	0.0023	0.0018	0.0027

Plot 248: (Kalama)

Series: (n248$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	-0.2882	0.4644	0.2197	0.1431	0.7135	0.6322
s.e.	0.5639	0.2785	0.1761	0.1415	0.5555	0.1561

Plot 251: (Kathiani)

Series: (n251$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	-0.2841	0.5217	0.1974	0.0994	0.6796	0.7089
s.e.	0.4423	0.2240	0.1902	0.1422	0.4275	0.1324

Plot 252: (Mwala)

Series: (n252$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	-0.346	0.5114	0.1979	0.1172	0.7638	0.7931
s.e.	0.352	0.1973	0.1624	0.1393	0.3256	0.0931

Plot 263: (Kangundo)

Series: (n263$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	-0.2427	0.5441	0.1827	0.0427	0.6205	0.7728
s.e.	0.0474	NaN	0.1023	0.0094	NaN	NaN

Plot 264: (Ndithini)

Series: (n264$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	-0.5126	0.7300	0.3112	-0.0812	1.000	0.7409
s.e.	0.1357	0.1643	0.1395	0.1483	0.026	0.1071

Plot 265: (Mavoko)

Series: (n265$rain)

ARIMA(4,0,1)(1,0,0)[36] with zero mean

Coefficients:

	ar1	ar2	ar3	ar4	ma1	sar1
	0.0463	0.3227	0.1777	0.1123	0.2800	0.5442
s.e.	0.8478	0.3168	0.2503	0.1712	0.8427	0.19